



Diffuse source separation in CMB observations

J. Delabrouille, J.-F. Cardoso

► To cite this version:

J. Delabrouille, J.-F. Cardoso. Diffuse source separation in CMB observations. École thématique. International summer school 'Data Analysis in Cosmology', 6-10 september 2005, Valencia, Spain, 2007, pp.47. cel-00162531

HAL Id: cel-00162531

<https://cel.hal.science/cel-00162531>

Submitted on 13 Jul 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Diffuse source separation in CMB observations

J. Delabrouille¹ and J.-F. Cardoso²

¹ APC, 11 place Marcelin Berthelot, F75231 Paris Cedex 05
delabrouille@apc.univ-paris7.fr

² LTCI, 46 rue Barrault, F75634 Paris Cedex 13 and
APC, 11 place Marcelin Berthelot, F75231 Paris Cedex 05
cardoso@enst.fr

1 Introduction

Spectacular advances in the understanding of the Big-Bang model of cosmology have been due to increasingly accurate observations of the properties of the Cosmic Microwave Background (CMB). The detector sensitivities of modern experiments have permitted to measure fluctuations of the CMB temperature with such a sensitivity that the contamination of the data by astrophysical foreground radiations, rather than by instrumental noise, is becoming the major source of limitation. This will be the case, in particular, for the upcoming observations by the Planck mission, to be launched by ESA in 2008 [Lamarre et al. (2003), Mandolesi et al. (2000), Lamarre et al. (2000)], as well as for next generation instruments dedicated to the observation of CMB polarisation.

In this context, the development of data analysis methods dedicated to identifying and separating foreground contamination from CMB observations is of the utmost importance for future CMB experiments. In many astrophysical observations indeed, and in particular in the context of CMB experiments, signals and images contain contributions from several components or *sources*. Some of these sources may be of particular interest (CMB or other astrophysical emission), some may be unwanted (noise).

Obviously, components cannot be properly studied in data sets in which they appear only as a mixture. Component separation consists, for each of them, in isolating the emission from all the other components present in the data, in the best possible way.

It should be noted that what “best” means depends on what the isolated data will be used for. Very often, one tries to obtain, for each component, an estimated map (or a set of maps at different frequencies) minimising the total error variance, i.e. minimising

$$\chi^2 = \sum_p |\hat{s}(p) - s(p)|^2 \quad (1)$$

where $s(p)$ is the true component emission, and $\hat{s}(p)$ its estimated value. p indexes the space of interest for the component, typically a set of pixels

(θ_p, ϕ_p) , or modes (ℓ, m) of a spherical harmonic decomposition of a full sky map, or a set of Fourier modes (k_x, k_y) ...

More generally, the objective of component separation is to estimate a *set of parameters* which describes the component of interest. In the simplest case, this set of parameters may be emission in pixels, but it may also be instead parameters describing statistical properties such as power spectra, spectral indices, etc... Since the set of parameters depends on the model assumed for the components, this model is of the utmost importance for efficient component separation. In the following, a significant part of the discussion will thus be dedicated to a summary of existing knowledge and of component modeling issues.

In the following, it is assumed that we are given a set of observations $y_i(p)$, where i , ranging from 1 to N_{chann} , indexes the observation frequency. The observed emission in each of the frequency bands is supposed to result from a mixture of several astrophysical components, with additional instrumental noise.

In this review paper, we discuss in some detail the problem of diffuse component separation. The paper is organised as follows: in the next section, we review the principles and implementation of the ILC, a very simple method to average the measurements obtained at different frequencies; section 3 reviews the known properties of diffuse sky emissions, useful to model the observations and put priors on component parameters; section 4 discusses observation and data reduction strategies to minimize the impact of foregrounds on CMB measurements, based on physical assumptions about the various emissions; section 5 discusses the model of a linear mixture, and various options for its linear inversion to separate astrophysical components; section 6 discusses a non linear solution for inverting a linear mixture, based on a maximum entropy method; section 7 presents the general ideas behind Blind Source Separation (BSS) and Independent Component Analysis (ICA); section 8 discusses a particular method, the Spectral Matching ICA (SMICA); section 10 concludes with a summary, hints about recent and future developments, open issues, and a prospective.

Let the reader be warned beforehand that this review may not do full justice to much of the work having been done on this very exciting topic. The discussion may be somewhat partial, although not intentionally. It has not been possible to the authors to review completely and compare all of the relevant work, partly for lack of time, and partly for lack of details in the published papers. As much as possible, we have nonetheless tried to mention all of the existing work, to comment the ideas behind the methods, and to quote most of the interesting and classical papers.

2 ILC: Internal Linear Combination

The Internal Linear Combination (ILC) component separation method assumes very little about the components. One of them (e.g. the CMB) is considered to be the only emission of interest, all the other being unwanted foregrounds.

It is assumed that the template of emission of the component of interest is the same at all frequencies of observation, and that the observations are calibrated with respect to this component, so that for each frequency channel i we have:

$$y_i(p) = s(p) + f_i(p) + n_i(p) \quad (2)$$

where $f_i(p)$ and $n_i(p)$ are foregrounds and noise contributions respectively in channel i .

A very natural idea, since all the observations actually measure $s(p)$ with some error $f_i(p) + n_i(p)$, consists in averaging all these measurements, giving a specific weight w_i to each of them. Then, we look for a solution of the form:

$$\hat{s}(p) = \sum_i w_i(p) y_i(p) \quad (3)$$

where the weights $w_i(p)$ are chosen to maximize some criterion about the reconstructed estimate $\hat{s}(p)$ of $s(p)$, while keeping the component of interest unchanged. This requires in particular that for all p , the sum of the coefficients $w_i(p)$ should be equal to 1.

2.1 Simple ILC

The simplest version of the ILC consists in minimising the variance σ^2 of the map $\hat{s}(p)$ using weights independent of p (so that $w_i(p) = w_i$ independent of p), with $\sum w_i = 1$. In this case, the estimated component is

$$\begin{aligned} \hat{s}(p) &= \sum_i w_i y_i(p) \\ &= s(p) + \sum_i w_i f_i(p) + \sum_i w_i n_i(p). \end{aligned} \quad (4)$$

Hence, under the assumption of de-correlation between $s(p)$ and all foregrounds, and between $s(p)$ and all noises, the variance of the error is minimum when the variance of the ILC map itself is minimum.

2.2 ILC implementation

We now outline a practical implementation of the ILC method. For definiteness (and simplicity), we will assume here that the data is in the form of harmonic coefficients $s(\ell, m)$. The variance of the ILC map is:

$$\sigma^2 = \sum_{\ell \geq 1} \sum_{m=-\ell}^{\ell} \mathbf{w}^\dagger \mathbf{C}(\ell, m) \mathbf{w} = \mathbf{w}^\dagger \mathbf{C} \mathbf{w} \quad (5)$$

where $\mathbf{C}(\ell, m) = \langle \mathbf{y}(\ell, m) \mathbf{y}^\dagger(\ell, m) \rangle$ is the covariance matrix of the observations in mode (ℓ, m) , and \mathbf{C} is the covariance summed over all modes except $\ell = 0$. $\mathbf{y}(\ell, m)$ and \mathbf{w} stand for the vectors of generic element $y_i(\ell, m)$ and w_i respectively. The minimum, under the constraint of $\sum w_i = 1$, is obtained, using the Lagrange multiplier method, by solving the linear system

$$\begin{aligned} \forall i, \quad \frac{\partial}{\partial w_i} \left[\sigma^2 + \lambda \left(1 - \sum w_i \right) \right] &= 0 \\ \sum_i w_i &= 1 \end{aligned} \quad (6)$$

Straightforward linear algebra gives the solution

$$w_i = \frac{\sum_j [\mathbf{C}^{-1}]_{ij}}{\sum_{ij} [\mathbf{C}^{-1}]_{ij}} \quad (7)$$

Note that if the template of emission of the component of interest is the same at all frequencies of observation, but the observations are *not* calibrated with respect to this component, equation 2 becomes:

$$y_i(p) = a_i s(p) + f_i(p) + n_i(p) \quad (8)$$

In this case, it is still possible to implement an ILC. The solution is

$$\mathbf{w} = \frac{\mathbf{A}^T \mathbf{C}^{-1}}{\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A}} \quad (9)$$

where \mathbf{A} is the vector of recalibration coefficients a_i . This solution of equation 9 is equivalent to first changing the units in all the observations to make the response 1 in all channels, and then implementing the solution of equation 7.

2.3 Examples of ILC separation: particular cases

This idea of ILC is quite natural. It has, however, several unpleasant features, which makes it non-optimal in most real-case situations. Before discussing this, let us examine now what happens in two simple particular cases.

Case 1: Noisy observations with no foreground

If there are no foregrounds, and the observations are simply noisy maps of $s(p)$, with independent noise for all channels of observation, the ILC solution should lead to a noise-weighted average of the measurements.

Let us assume for simplicity that we have two noisy observations, y_1 and y_2 , with $y_i = s + n_i$. In the limit of very large maps, so that cross correlations

between s , n_1 and n_2 vanish, the covariance matrix of the observations takes the form:

$$\mathbf{C} = \begin{bmatrix} S + N_1 & S \\ S & S + N_2 \end{bmatrix}$$

where S is the variance of the signal (map) of interest, and N_1 and N_2 the noise variances for the two channels. The inverse of \mathbf{C} is:

$$\mathbf{C}^{-1} = \frac{1}{\det(\mathbf{C})} \begin{bmatrix} S + N_2 & -S \\ -S & S + N_1 \end{bmatrix}$$

and applying equation 7, we get $w_1 = N_2/(N_1 + N_2)$ and $w_2 = N_1/(N_1 + N_2)$. This is the same solution as weighting each map i proportionally to $1/N_i$.

Case 2: Noiseless observations with foregrounds

Let us now examine the opposite extreme, where observations are noiseless linear mixtures of several astrophysical components. Consider the case of two components, with two observations. We can write the observations as $\mathbf{y} = \mathbf{A}\mathbf{s}$, where \mathbf{A} is the so-called “mixing matrix”, and $\mathbf{s} = (s_1, s_2)^\dagger$ the vector of sources.

The covariance of the observations is

$$\mathbf{C} = \mathbf{y}\mathbf{y}^\dagger = \mathbf{A}\mathbf{s}\mathbf{s}^\dagger\mathbf{A}^\dagger$$

and its inverse is

$$\mathbf{C}^{-1} = [\mathbf{A}^\dagger]^{-1} [\mathbf{s}\mathbf{s}^\dagger]^{-1} \mathbf{A}^{-1} \quad (10)$$

Let us assume that we are interested in the first source. The data are then calibrated so that the mixing matrix \mathbf{A} and its inverse are of the form

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} \\ 1 & a_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} a_{22} & -a_{12} \\ -1 & 1 \end{bmatrix}$$

Then, if we assume that components 1 and 2 are uncorrelated, equation 10 yields

$$\mathbf{C}^{-1} = \frac{1}{(\det(\mathbf{A}))^2} \begin{bmatrix} a_{22} & -1 \\ -a_{12} & 1 \end{bmatrix} \begin{bmatrix} S_1^{-1} & 0 \\ 0 & S_2^{-1} \end{bmatrix} \begin{bmatrix} a_{22} & -a_{12} \\ -1 & 1 \end{bmatrix} \quad (11)$$

where S_1 and S_2 are the variances of components 1 and 2 respectively. After expansion of the matrix product, we get:

$$\mathbf{C}^{-1} = \frac{1}{(\det(\mathbf{A}))^2} \begin{bmatrix} (a_{22}^2 S_1^{-1} + S_2^{-1}) & (-a_{22} a_{12} S_1^{-1} - S_2^{-1}) \\ (-a_{22} a_{12} S_1^{-1} - S_2^{-1}) & (a_{12}^2 S_1^{-1} + S_2^{-1}) \end{bmatrix} \quad (12)$$

and using equation 7, we get

$$\mathbf{w} = \frac{1}{(a_{22} - a_{12})} \begin{bmatrix} a_{22} \\ -a_{12} \end{bmatrix} \quad (13)$$

which is the transpose of the first line of matrix \mathbf{A}^{-1} , so that $\hat{s}_1 = \mathbf{w} \cdot \mathbf{y} = s_1$. As expected, *if the covariance of the two components vanishes*, the ILC solution is equivalent, for the component of interest, to what is obtained by inversion of the mixing matrix.

What happens now if the two components are correlated? Instead of the diagonal form $\text{diag}(S_1, S_2)$, the covariance matrix of the sources contains an off-diagonal term S_{12} , so that equation 11 becomes:

$$\mathbf{C}^{-1} = \frac{1}{(\det(\mathbf{A}))^2} \frac{1}{(S_1 S_2 - S_{12}^2)} \begin{bmatrix} a_{22} & -1 \\ -a_{12} & 1 \end{bmatrix} \begin{bmatrix} S_2 & S_{12} \\ S_{12} & S_1 \end{bmatrix} \begin{bmatrix} a_{22} - a_{12} \\ -1 & 1 \end{bmatrix} \quad (14)$$

which yields the solution

$$\mathbf{w} = \frac{1}{(a_{22} - a_{12})} \begin{bmatrix} a_{22} + S_{12}/S_2 \\ -a_{12} - S_{12}/S_2 \end{bmatrix} \quad (15)$$

The ILC is not equivalent anymore to the inversion of the mixing matrix \mathbf{A} . Instead, the estimate \hat{s}_1 of s_1 is:

$$\hat{s}_1 = \mathbf{w} \cdot \mathbf{y} = s_1 - \frac{S_{12}}{S_2} s_2 \quad (16)$$

The ILC result is biased, giving a solution in which a fraction of s_2 is subtracted erroneously, in proportion to the correlation between s_1 and s_2 , to lower as much as possible the variance of the output map. The implication of this is discussed in paragraph 2.6.

2.4 Improving the ILC method

With the exception of the CMB, diffuse sky emissions are known to be very non stationnary (e.g. galactic foregrounds are strongly concentrated in the galactic plane). In addition, most of the power is concentrated on large scales (the emissions are strongly correlated spatially). As the ILC method minimizes the total variance of the ILC map (the integrated power from all scales, as can be seen in equation 5), the weights w_i are strongly constrained essentially by regions of the sky close to the galactic plane, where the emission is strong, and by large scales, which contain most of the power. In addition, the ILC method finds weights resulting from a compromise between reducing astrophysical foreground contamination, and reducing the noise contribution. In other words, for a smaller variance of the output map, it pays off more to reduce the galactic contamination in the galactic plane and on large scales, where it is strong, rather than at high galactic latitude and on small scales, where there is little power anyway. This particularity of the ILC when implemented globally is quite annoying for CMB studies, for which all scales are interesting, and essentially the high galactic latitude data is useful.

Away from the galactic plane and on small scales, the best linear combination for cleaning the CMB from foregrounds and noise may be very different

from what it is close to the galactic plane and on large scales. A very natural idea to improve on the ILC is to decompose sky maps in several regions and/or scales, and apply an ILC independently to all these maps. The final map is obtained by adding-up all the ILC maps obtained independently in various regions and at different scales. Applications of these ideas are discussed in the next paragraph.

2.5 ILC-based foreground-cleaned CMB map from WMAP data

A map of CMB anisotropies has been obtained using the ILC method [Bennett et al. (2003)] on first year data from the WMAP mission, and has been released to the scientific community as part of the first year WMAP data products.

The input data is the set of five all sky, band averaged maps for the K, Ka, Q, V and W frequency bands, all of which smoothed to the same 1 degree resolution for convenience. The ILC is performed independently in 12 regions, 11 of which being in the WMAP kp2 mask at low galactic latitudes, designed to mask out regions of the sky highly contaminated by galactic foregrounds. This division into twelve regions is justified by the poor performance of the ILC on the full sky, interpreted as due to varying spectral indices of the astrophysical foregrounds. Discontinuities between the regions are reduced by using smooth transitions between the regions.

Little detail is provided on the actual implementation of the ILC by the WMAP team. Apparently, a non-linear iterative minimization algorithm was used, instead of the linear solution outlined in paragraph 2.2. Although there does not seem to be any particular reason for this choice, in principle the particular method chosen to minimize the variance does not matter, as long as it finds the minimum efficiently. There seem to be, however, indications that the convergence was not perfect, as discussed by Eriksen and collaborators in a paper discussing the ILC and comparing the results of the several of its implementations on WMAP data [Eriksen et al. (2004)]. Caution should probably be taken when using the WMAP ILC map for any purpose other than a visual impression of the CMB.

Tegmark and collaborators have improved the ILC method in several respects, and provide an independent CMB map obtained from WMAP data by ILC [Tegmark et al. (2003)]. Their implementation allows the weights to depend not only on the region of the sky, but also on angular scale, as discussed in paragraph 2.4. In addition, they partially deconvolve the WMAP maps in harmonic space to put them all to the angular resolution of the channel with the smallest beam, rather than smoothing all maps to put them all to the angular resolution of the channel with the largest beam. As a result, their ILC map has better angular resolution, but higher total noise. The high resolution map, however, can be filtered using a Wiener filter for minimal variance of the error. The Wiener-filtered map is obtained by multiplying each $a_{\ell m}$ mode of the map by a factor

$$W(\ell, m) = C_\ell / S_\ell$$

where C_ℓ is the estimated CMB power spectrum (computed for the cosmological model fitting best the WMAP data estimate), and S_ℓ is the estimated power spectrum of the noisy CMB map obtained by the authors using their ILC method.

The CMB map obtained by the WMAP team from first year data is shown in figure 1. For comparison, the map obtained by Tegmark et al. is shown in figure 2. Both give a good visual perception of what the CMB field looks like.

2.6 Comments about the ILC

The ILC has been used essentially to obtain a clean map of CMB emission. In principle, nothing prevents using it also for obtaining cleaned maps of other emissions, with the caveats that

- The data must be calibrated with respect to the emission of interest, so that the data takes the form of equation 2. This implies that the template of emission of the component of interest should not change significantly with the frequency-band of observation. This is the case for the CMB (temperature and polarisation), or for the SZ effect (to first order at least... more on this later).
- The component of interest should not be correlated with other components. Galactic components, being all strongly concentrated in the galactic plane, can thus not be recovered reliably with the ILC.

This issue of decorrelation of the component of interest $s(p)$ and the foregrounds can also generate problems in cases where the empirical correlation between the components does not vanish. As demonstrated in 2.3, the ILC method will not work properly (biasing the result) if the assumption the component of interest $s(p)$ are correlated for whatever reason. In particular, small data sets, even if they are realisations of actually uncorrelated random processes, are always empirically correlated to some level. For this reason, the ILC should not be implemented independently on too small subsets of the original data (very small regions, very few modes).

Finally, whereas the ILC is a powerful tool when nothing is known about the data, it is certainly non optimal when prior information is available. Foreground emissions are discussed in some detail in following section.

3 Sky emission model: components

“Know your enemy”... This statement, borrowed from elementary military wisdom, applies equally well in the fight against foreground contamination. Prior knowledge about astrophysical components indeed has been widely used in all practical CMB data analyses. Methods can then be specifically tailored

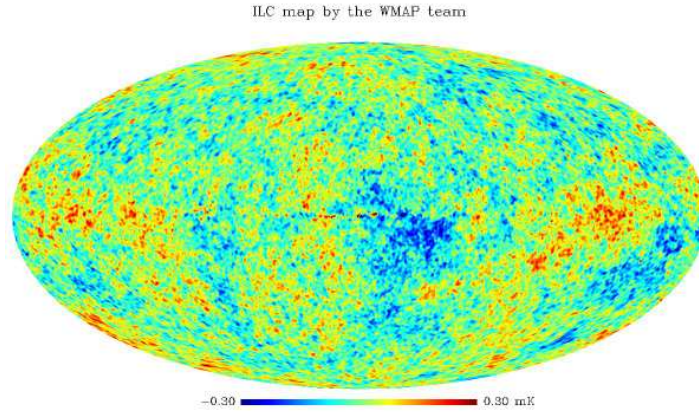


Fig. 1. The ILC map of the CMB obtained by the WMAP team (one year data). Residuals of galactic emission are clearly visible in the center of the map. The color scale spans a range of -300 to $+300$ μK thermodynamic, although localised residuals exceed these values.

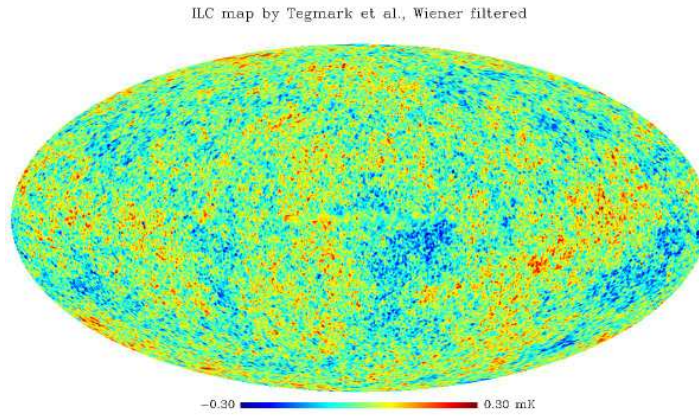


Fig. 2. The foreground-cleaned CMB map of Tegmark et al., obtained by the ILC method described in [Tegmark et al. (2003)], after Wiener filtering. The effect of the region and scale-dependent weighting can be seen in the center of the map (galactic center) where the map looks smoother and flatter than elsewhere. The color scale spans a range of -300 to $+300$ μK , although localised residuals exceed these values, as in figure 1. The superior angular resolution can clearly be seen.

to remove foregrounds based on their physical properties, in particular their morphology, their localisation, and their frequency scaling based on the physical understanding of their emission mechanisms.

In addition to knowledge about the unwanted foregrounds, prior knowledge about the component of interest is of the utmost importance for its identification and separation in observations. In the ILC method discussed above, for instance, the prior knowledge of the emission law of the CMB (derivative of a blackbody) is specifically used.

3.1 The various astrophysical emissions

Astrophysical emissions relevant to the framework of CMB observations can be classified in three large categories (in addition to the CMB itself). Diffuse galactic emission, extragalactic emission, and solar system emission.

Diffuse galactic emissions originate from the local interstellar medium (ISM) in our own galaxy. The ISM is constituted of cold clouds of molecular or atomic gas, of an intercloud medium which can be partly ionised, and of hot ionized regions presumably formed by supernovae. These different media are strongly concentrated in the galactic plane. The intensity of corresponding emissions decreases with galactic latitude with a cosecant law behaviour (the optical depth of the emitting material scales proportionally to $1/\sin b$). Energetic free electrons spiralling in the galactic magnetic field generate synchrotron emission, which is the major foreground at low frequencies (below a few tens of GHz). Warm ionised material emits free-free (Bremstrahlung) emission, due to the interaction of free electrons with positively charged nuclei. Small particles of matter (dust grains and macromolecules) emit radiation as well, through thermal greybody emission, and possibly through other mechanisms.

Extragalactic emissions arise from a large background of resolved and unresolved radio and infrared galaxies, as well as clusters of galaxies. The thermal and kinetic Sunyaev-Zel'dovich effects, due to the inverse Compton scattering of CMB photons off hot electron gas in ionized media, are of special interest for cosmology. These effects occur, in particular, towards clusters of galaxies, which are known to comprise a hot (few keV) electron gas. Infrared and radiogalaxies emit also significant radiation in the frequency domain of interest for CMB observations, and contribute both point source emission from nearby bright objects, and a diffuse background due to the integrated emission of a large number of unresolved sources, too faint to be detected individually, but which contribute sky background inhomogeneities which may pollute CMB observations.

Solar system emission comprises emissions from the planets, their satellites, and a large number of small objects (asteroids). In addition to those, there is diffuse emission due to dust particles and grains in the ecliptic plane (zodiacal light). The latter is significant essentially at the highest frequencies of an instrument like the Planck HFI [Maris et al. (2006)].

In the rest of this section, we briefly outline the general properties of these components and the modeling of their emission in the centimetre to sub-millimetre wavelength range.

3.2 The Cosmic Microwave Background

The cosmic microwave background, relic radiation from the hot big bang emitted at the time of decoupling when the Universe was about 370,000 years old, is usually thought of (by cosmologists) as the component of interest in the sky emission mixture. Millimetre and submillimetre wave observations, however, sometimes aim not only at measuring CMB anisotropies, but also other emissions. In this case, the CMB becomes a noxious background which has to be subtracted out of the observations, just as any other.

The CMB emission is relatively well known already. The main theoretical framework of CMB emission can be found in any modern textbook on cosmology, as well as in several reviews [Hu & Dodelson (2002), White & Cohn (2002)]. The achievement of local thermal equilibrium in the primordial plasma before decoupling, together with the very low level of the perturbations, guaranties that CMB anisotropies are properly described as the product of a spatial template $\Delta T(p) = T_{\text{CMB}}(p) - \bar{T}_{\text{CMB}}$, and a function of ν (frequency scaling) which is the derivative of a blackbody with respect to temperature:

$$\Delta I_\nu(p) = \Delta T_{\text{CMB}}(p) \left[\frac{\partial B_\nu(T)}{\partial T} \right]_{T=\bar{T}_{\text{CMB}} \simeq 2.726 \text{ K}} \quad (17)$$

In the standard cosmological model, the CMB temperature fluctuation map $\Delta T(p)$ is expected to be a realisation of a stationary Gaussian random field, with a power spectrum C_ℓ displaying a series of peaks and troughs (the acoustic peaks), the location and relative size of which are determined by a few free parameters of the cosmological model.³

Good maps of sky emission at a resolution of about 15 arcminutes, obtained from WMAP data in the frequency range 20–90 GHz, clearly comprise at high galactic latitude an astrophysical component compatible with all these predictions. The power spectrum is measured with excellent accuracy by WMAP up to the second Doppler peak, while complementary balloon-borne and ground-based experiments yield additional measurements at higher ℓ (smaller scales).

Efficient diffuse component separation methods should make use of this current status of knowledge about the CMB:

³ The power spectrum C_ℓ is defined as the set of variances of the coefficients $a_{\ell m}$ of the expansion of the random field representing CMB relative temperature fluctuations $\Delta T(p)/\bar{T}_{\text{CMB}}$ onto the basis of spherical harmonics on the sphere $Y_{\ell m}(\theta, \phi)$. The stationarity and isotropy of the random field guarantees that the variance of $a_{\ell m}$ (coefficients C_ℓ) is independent of m .

- Law of emission, known to a high level of precision to be the derivative of a blackbody with respect to temperature, as expected theoretically and checked experimentally with the Boomerang [de Bernardis et al. (2000)] and Archeops [Tristram et al. (2005)] multifrequency data sets, as well as with the WMAP data [Bennett et al. (2003), Patanchon et al. (2005)]
- Stationarity and gaussianity to a high level of accuracy, as expected theoretically and checked on WMAP data [Komatsu et al. (2003)]
- Good cosmological prior on the power spectrum of the fluctuations, validated experimentally with several data sets [Netterfield et al. (2002), Hinshaw et al. (2006)]

A good visual impression of all-sky CMB emission is given in figures 1 and 2. The present status of knowledge of the power spectrum C_ℓ is shown in figure 3.

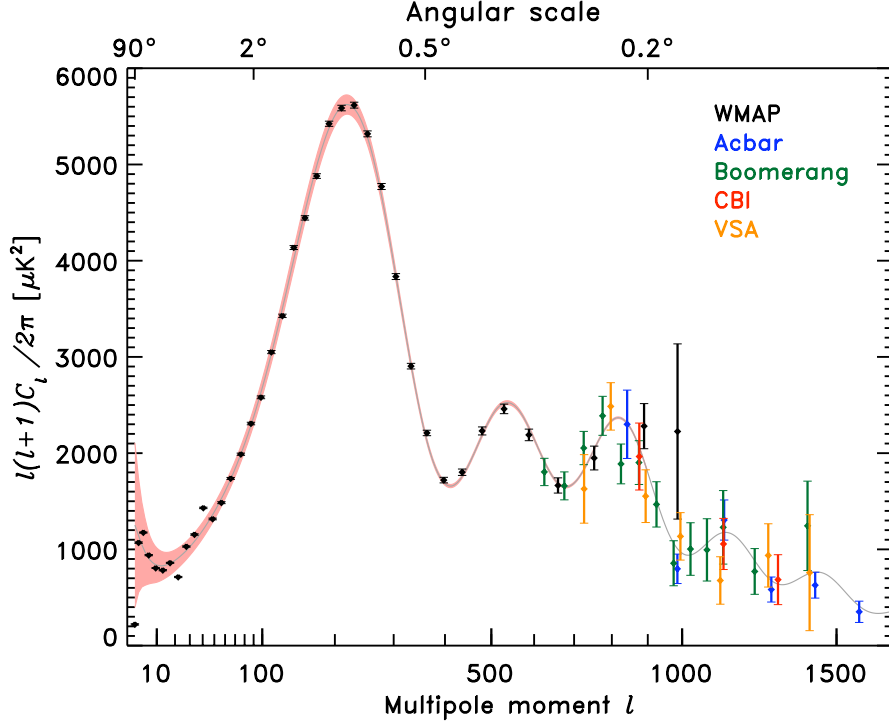


Fig. 3. Present-day best constraints of the CMB temperature power spectrum (from [Hinshaw et al. (2006)]). Data sets in addition to WMAP 3-year data are from [Jones et al. (2005), Kuo et al. (2004), Readhead et al. (2004), Dickinson et al. (2004)]

The extraction of CMB emission from a set of multifrequency observations may be done with the following objectives in mind (at least):

- Get the best possible map of the CMB (in terms of total least square error, from noise and foregrounds together);
- Get the CMB map with the least possible foreground contamination;
- Get the CMB map for which spurious non-gaussianity from foregrounds, noise and systematic effects is minimal;
- Get the best possible estimate of the CMB angular power spectrum...

Obviously, the best component separation method for extracting the CMB will depend on which of the above is the primary objective of component separation.

3.3 Emissions from the diffuse interstellar medium

Synchrotron emission

Synchrotron emission arises from energetic charged particles spiralling in a magnetic field. In our galaxy, such magnetic fields extend outside the galactic plane. Energetic electrons originating from supernovae shocks, spiraling in this field, can depart the galactic plane and generate emission even at high galactic latitudes. For this reason, synchrotron emission is less concentrated in the galactic plane than free-free and dust.

The frequency scaling of synchrotron emission is a function of the distribution of the energies of the radiating electrons. For number density distributions $N(E) \propto E^{-\gamma}$, the flux emission is also in the form of a power law, $I_\nu \propto \nu^{-\alpha}$, with $\alpha = (\gamma - 1)/2$. In Rayleigh-Jeans (RJ) temperature $\Delta T \propto \nu^{-\beta}$ with $\beta = \alpha + 2$. Typically, β ranges from 2.5 to 3.1, and is somewhat variable across the sky.

In spite of a moderate sensitivity for current standards, the 408 MHz all sky map [Haslam et al. (1981)], dominated by synchrotron emission, gives a good visual impression of the distribution of synchrotron over the sky.

In principle, synchrotron emission can be highly polarised, up to 50-70%

Free-Free emission

Free-free emission is the least well known observationally of the three major emissions originating from the galactic interstellar medium in the millimetre and centimetre wavelength range. This emission arises from the interaction of free electrons with ions in ionised media, and is called “free-free” because of the unbound state of the incoming and outgoing electron. Alternatively, free-free is called “Bremsstrahlung” emission (“braking radiation” in German), because photons are emitted while electrons loose energy by interaction with the heavy ions.

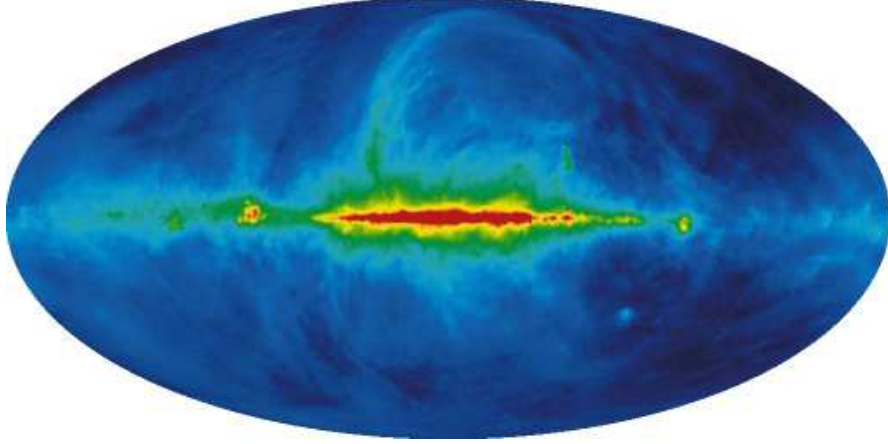


Fig. 4. The 408 MHz all-sky synchrotron map [Haslam et al. (1981)]. Data and images are available on the NASA Lambda web site.

Theoretical calculations of free-free emission in an electrically neutral medium consisting of ions and electrons gives an estimate of the brightness temperature at frequency ν for free-free emission of the form:

$$T_{\text{ff}} \simeq 0.08235 T_e^{-0.35} \nu^{-\beta} \int_{\text{l.o.s.}} N_e N_i dl \quad (18)$$

where T_e is in Kelvin, ν is in GHz and the line of sight integral of electron and ion density in cm^{-6}pc [Smoot (1998)]. Theoretical estimates of the spectral index, β , range from about 2.1 to 2.15, with errors of ± 0.03 .

While free-free emission is not observed directly, as it never dominates over other astrophysical emissions, the source of its emission (mainly ionised hydrogen clouds) can be traced with hydrogen recombination emission lines, and particularly $\text{H}\alpha$ emission. The connection between $\text{H}\alpha$ and free-free has been discussed extensively by a number of authors [Smoot (1998), Valls-Gabaud (1998), McCullough et al. (1999)]. We have:

$$\frac{T_{\text{ff}}[\text{mK}]}{I_{\alpha}[\text{R}]} \simeq 10.4 \nu^{-2.14} T_4^{0.527} 10^{0.029/T_4} (1 + 0.08) \quad (19)$$

Where $T_{\text{ff}}[\text{mK}]$ is the free-free brightness temperature in mK, $I_{\alpha}[\text{R}]$ the $\text{H}\alpha$ surface brightness in Rayleigh, ν the frequency, and T_4 the temperature of the ionized medium in units of 10^4 K. The Rayleigh (R) is defined as $1 \text{ R} = (10^6/4\pi) \text{ photons/cm}^2/\text{s}/\text{sr}$.

Free-free emission, being due to incoherent emissions from individual electrons scattered by nuclei in a partially ionised medium, is not polarised (to first order at least).

Thermal emission of galactic dust

The present knowledge of interstellar dust is based on extinction observations from the near infrared to the UV domain, and on observations of its emission from radio frequencies to the infrared domain.

Dust consists in small particles of various materials, essentially silicate and carbonaceous grains of various sizes and shapes, in amorphous or crystalline form, sometimes in aggregates or composites. Dust is thought to comprise also large molecules of polycyclic aromatic hydrocarbon (PAH). The sizes of the grains range from few nanometers for the smallest, to micrometers for the largest. They can emit through a variety of mechanisms. The most important for CMB observations is greybody emission in the far infrared, at wavelengths ranging from few hundreds of microns to few millimeters. The greybody emission is typically characterised by a temperature T_{dust} and by an emissivity proportional to a power of the frequency ν :

$$I_\nu \propto \nu^\beta B_\nu(T_{\text{dust}}) \quad (20)$$

where $B_\nu(T)$ is the usual blackbody emission

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \quad (21)$$

This law is essentially empirical. In practice, dust clouds along the line of sight can have different temperatures and different compositions: bigger or smaller grains, different materials. They can thus have different emissivities as well. Temperatures for interstellar dust are expected to range from about 5 Kelvin to more than 30 Kelvin, depending on the heating of the medium by radiation from nearby stars, with typical values of 16-18 K for emissivity indices $\beta \simeq 2$.

In principle, thermal emission from galactic dust should not be strongly polarised, unless dust particles are significantly asymmetric (oblate or prolate), and there exists an efficient process for aligning the dust grains in order to create a significant statistical asymmetry. Preliminary dust observations with the Archeops instrument [Benoît et al. (2004), Ponthieu et al. (2005)] seems to indicate polarisation levels of the order of few per cent, and as high as 15-20 per cent in a few specific regions.

Spinning dust or anomalous dust emission

In the last years, increasing evidence for dust-correlated emissions at frequencies below 30 GHz, in excess to expectations from synchrotron and free-free, has aroused interest in possible non-thermal emissions from galactic dust [Kogut et al. (1996), Leitch et al. (1997)]. Among the possible non-thermal emission mechanisms, spinning dust grains offer an interesting option [Draine & Lazarian (1998)].

At present, there still is some controversy on whether the evidence for non-thermal dust emission is robust enough for an unambiguous statement. Observations of different sky regions, indeed, yield somewhat different results [Dickinson et al. (2006), Fernández-Cerezo et al. (2006)], which may be due either to varying dust cloud properties, or to differences in the analyses and interpretations, or both. Certainly, investigating this question is an objective of primary interest for diffuse component separation methods (especially blind ones) in the near future.

3.4 The SZ effect

The Sunyaev Zel'dovich (SZ) effect [Sunyaev & Zeldovich (1972)] is the inverse Compton scattering of CMB photons on free electrons in ionised media. In this process, the electron gives a fraction of its energy to the scattered CMB photon. There are, in fact, several SZ effects: The thermal SZ effect is due to the scattering of photons on a high temperature electron gas, such as can be found in clusters of galaxies. The kinetic SZ effect is due to the scattering on a number of electrons with a global radial bulk motion with respect to the cosmic background. Finally, the polarised SZ effect is a second order effect due to the kinematic quadrupole of the CMB in the frame of an ensemble of electrons with a global transverse bulk motion with respect to the CMB.

SZ effects are not necessarily linked to clusters of galaxies. Any large body with hot ionised gas can generate significant effects. It has been proposed that signatures of inhomogeneous reionisation can be found via the kinetic and thermal SZ effect [Aghanim et al. (1996), Gruzinov & Hu (1998), Yamada et al. (1999)]. However, the largest expected SZ signatures originate from ionised intra-cluster medium.

Clusters of galaxies

Clusters of galaxies, the largest known massive structures in the Universe, form by gravitational collapse of matter density inhomogeneities on large scales (comoving scales of few Mpc). They can be detected either optically from concentrations of galaxies at the same redshift, or in the submillimeter by their thermal SZ emission, or by the effect of their gravitational mass in weak shear maps, or in X-ray. The hot intracluster baryonic gas can be observed through its X-ray emission due to Bremsstrahlung (free-free) emission of the electrons on the nuclei, which permits to measure the electron temperature (typically a few keV). On the sky, typical cluster angular sizes range from about one arcminute to about one degree. Clusters are scattered over the whole sky, although this distribution follows the repartition of structure on the largest scales in the universe. Large scale SZ effect observations may be also used to survey the distribution of hot gas on these very large scales, although such SZ emission, from filaments and pancakes in the distribution, is

expected to be at least an order of magnitude lower in intensity than thermal SZ emission from the clusters themselves.

Each cluster of galaxies has its own thermal, kinetic and polarised SZ emission. These various emissions and their impact on CMB observations and for cosmology have been studied by a variety of authors. Useful reviews have been made by Birkinshaw [Birkinshaw (1999)] and Rephaeli [Rephaeli (2002)], for instance.

Thermal SZ

The thermal SZ effect generated by a gas of electrons at temperature T_e is, in fact, a spectral distortion of the CMB emission law. It is common to consider as the effect the difference $\Delta I_\nu = I_\nu - B_\nu(T_{\text{CMB}})$ between the distorted CMB photon distribution I_ν and the original one $B_\nu(T_{\text{CMB}})$. In the non-relativistic limit (when T_e is lower than about 5 keV, which is the case for most clusters), the shape of the spectral distortion does not depend on the temperature. The change in intensity due to the thermal SZ effect is:

$$\Delta I_\nu = y \frac{x e^x}{(e^x - 1)} \left[\frac{x(e^x + 1)}{(e^x - 1)} - 4 \right] B_\nu(T_{\text{CMB}}) \quad (22)$$

where $B_\nu(T_{\text{CMB}})$ is the Planck blackbody emission law at CMB temperature

$$B_\nu(T_{\text{CMB}}) = \frac{2h\nu^3}{c^2} \frac{1}{e^x - 1}$$

and $x = h\nu/kT_{\text{CMB}}$. The dimensionless parameter y (Comptonisation parameter) is proportional to the integral of the electron pressure along the line of sight:

$$y = \int_{\text{los}} \frac{kT_e}{m_e c^2} n_e \sigma_{\text{thomson}} dl$$

where T_e is the electron temperature, m_e the electron mass, c the speed of light, n_e the electron density, and σ_{thomson} the Thomson cross section.

Kinetic SZ

The kinetic SZ effect is generated by the scattering of CMB photons off an electron gas in motion with respect to the CMB. This motion generates spectral distortions with the same frequency scaling as CMB temperature fluctuations, and are directly proportionnal to the velocity of the electrons along the line of sight. As the effect has the same frequency scaling as CMB temperature fluctuations, it is, in principle, indistinguishable from primordial CMB. However, since the effect arises in clusters of galaxies with typical sizes 1 arcminute, it can be distinguished to some level from the primordial CMB by spatial filtering, especially if the location of the clusters most likely to generate the effect is known from other information (e.g. the detection of the clusters through the thermal SZ effect).

Polarised SZ

The polarised SZ effect arises from the polarisation dependence of the Thomson cross section:

$$\sigma_T \propto |\mathbf{e}_1 \cdot \mathbf{e}_2|^2$$

where \mathbf{e}_1 and \mathbf{e}_2 are the polarisation states of the incoming and outgoing photon respectively. A quadrupole moment in the CMB radiation illuminating the cluster electron gas generates a net polarisation, at a level typically two orders of magnitude lower than the kinetic SZ effect [Sunyaev & Zeldovich (1980), Audit & Simmons (1999), Sazonov & Sunyaev (1999)]. Therefore, the kinetic SZ effect has been proposed as a probe to investigate the dependence of the CMB quadrupole with position in space. Cluster transverse motions at relativistic speed, however, generate also such an effect from the kinematic quadrupole induced by the motion. Multiple scattering of CMB photons also generates a low-level polarisation signal towards clusters.

The polarised SZ effects has a distinctive frequency scaling, independent (to first order) to cluster parameters and to the amplitude of the effect. Amplitudes are proportionnal:

- to τ for the intrinsic CMB quadrupole effect,
- to $(v_t/c)^2 \tau$ for the kinematic quadrupole effect
- to $(kT_e/m_e c^2) \tau^2$ and $(v_t/c) \tau^2$ for polarisation effects due to double scattering.

Here τ is the optical depth, v_t the transverse velocity, c the speed of light, k the Boltzmann constant, and T_e and m_e the electron temperature and mass.

As polarised effects arise essentially in galaxy clusters, they can be sought essentially in places where the much stronger thermal effect is detected, which will permit to improve the detection capability significantly. Polarised SZ emission, however, is weak enough that it is not expected to impact significantly the observation of any of the main polarised emissions.

Diffuse component or point source methods for SZ effect separation?

The SZ effect is particular in several respects. As most of the emission comes from compact regions towards clusters of galaxies (at arcminute scales), most of the present-day CMB experiments do not resolve clusters individually (apart for a few known extended clusters). For this reason, it seems natural to use point source detection methods for cluster detection (see review by Barreiro [Barreiro (2005)]). However, the very specific spectral signature, the presence of a possibly large background of clusters with emission too weak for individual cluster detection, and the interesting possibility to detect larger scale diffuse SZ emission, makes looking for SZ effect with diffuse component separation methods an interesting option.

3.5 The infrared background of unresolved sources

The added emissions from numerous unresolved infrared sources at high redshift make a diffuse infrared background, detected originally in the FIRAS and DIRBE data [Puget et al. (1996)]. Because each source has its specific emission law, and because this emission law is redshifted by the cosmological expansion, the background does not have a very well defined frequency scaling. It appears thus, in the observations at various frequencies, as an excess emission correlated between channels. The fluctuations of this background are expected to be significant at high galactic latitudes (where not masked by much stronger emissions from our own galaxy), and essentially at high frequencies (in the highest frequency channels of the Planck HFI).

3.6 Point sources

The “point sources” component comprises all emissions from astrophysical objects such as radio galaxies, infrared galaxies, quasars, which are not resolved by the instruments used in CMB observations. For such sources, the issues are both their detection and the estimation of parameters describing them (flux at various frequencies, location, polarisation...), and specific methods are devised for this purpose. For diffuse component separation, they constitute a source of trouble. Usually, pixels contaminated by significant point source emission are blanked for diffuse component separation.

4 Reduction of foreground contamination

The simplest way of avoiding foreground contamination consists in using prior information on emissions to reduce their impact on the data: by adequate selection of the region of observation, by masking some directions in the sky, by choosing the frequency bands of the instrument, or, finally, by subtracting an estimate of the contamination. All of these methods have been used widely in the context of CMB experiments.

4.1 Selection of the region of observation

Perhaps the most obvious solution to avoid contamination by foregrounds is to design the observations in such a way that the contamination is minimal. This sensible strategy has been adopted by ground-based and balloon-borne experiments observing only a part of the sky. In this case, CMB observations are made away from the galactic plane, in regions where foreground contamination from the galactic emissions is known to be small. The actual choice of regions of observation may be based on existing observations of dust and synchrotron emission at higher and lower frequencies, picking those regions where the emission of these foregrounds is known to be the lowest.

The drawback of this strategy is that the observations do not permit to estimate very well the level of contamination, nor the properties of the foregrounds.

4.2 Masking

For all-sky experiments, a strategy for keeping the contamination of CMB observations by foregrounds consists in masking regions suspected to comprise significant foreground emissions, and deriving CMB properties (in particular the CMB power spectrum) in the “clean” region. The drawback of this strategy is that sky maps are incomplete.

Typically, for CMB observations, pixels contaminated by strong point sources (radio and infrared galaxies) are blanked, as well as a region containing the galactic plane. Such masks have been used in the analysis of WMAP data.

4.3 Selection of the frequency bands of the instrument

Of course, the selection of the frequency of observation to minimize the overall foreground contamination is a sensible option. For this reason, many CMB experiments aim at observing the sky around 70–100 GHz. Ground-based observations, however, need to take into account the additional foreground of atmospheric emission, which leaves as best windows frequency bands around 30 GHz, 90 GHz, 150 GHz, and 240 GHz.

Figure 5 shows the expected typical frequency scalings for the major diffuse emission astrophysical components, including the CMB. For efficient component separation, CMB experiment, ideally, should comprise two or three channels around 70–100 GHz where CMB dominates, one channel around 217 GHz (the zero of the SZ effect), two channels at higher frequencies to monitor dust emission, and 3–4 channels at lower frequencies to monitor low frequency foregrounds.

Below 100 GHz, the present state of the art technology suggests the use of radiometers with high electron mobility transistor (HEMT) amplifiers, whereas above 100 GHz, low temperature bolometers provide a significantly better sensitivity than any other techniques. Typically, a single experiment uses one technology only. For Planck specifically, two different instruments have been designed to cover all the frequency range from 30 to 850 GHz.

4.4 Foreground cleaning

As a refinement to the above simple observational strategies, a first-order estimate of foreground contamination, based on observations made at low and high frequencies, can be subtracted from the observations. Depending on the accuracy of the model, the overall level of contamination can be reduced

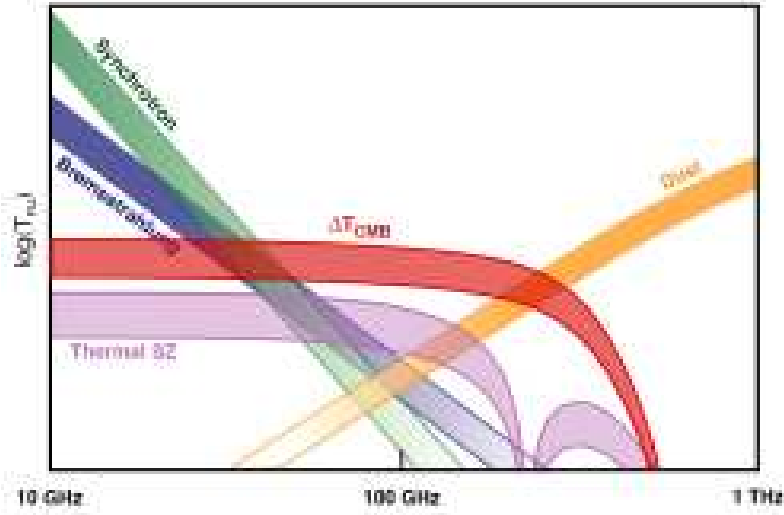


Fig. 5. The frequency scaling of CMB and most relevant diffuse foregrounds, in Rayleigh-Jeans temperature, between 10 GHz and 1 THz. Depending on the relative amplitude of synchrotron, bremsstrahlung and dust emissions, the minimum of galactic foregrounds is somewhat below 100 GHz. Free-free emission decreases roughly as $\nu^{-2.1}$ and synchrotron as ν^{-3} , while dust increases as ν^2 . The SZ effect is the major emission towards rich clusters, but is very localised. The thickness of the bands illustrates uncertainties as to the level of foregrounds, as well as uncertainties in the frequency scaling for synchrotron, free-free and dust emissions. Anomalous dust emission is not represented, due to our present lack of knowledge of the existence and nature of such a component.

by a factor of a few at least, which permits to reduce the amount of cut sky. This strategy, in particular, has been used by the WMAP team for the analysis of first year WMAP data [Bennett et al. (2003)].

Observations at low frequencies (10-40 GHz) can be used to map synchrotron emission and model its contribution in the 70-100 GHz range. Similar strategies can be used towards the high frequency side to model dust emission and subtract its contribution from CMB channels. For this purpose, models of emission as good as possible are needed, and the cleaning can be no better than the model used. There is always, therefore, a trade-off between a sophisticated model with simple correction methods (subtraction of an interpolation, simple decorrelation), and a simple model with sophisticated statistical treatments (multi-frequency filtering, independent component analysis). Which approach is best depends on a number of issues, and the answer is not completely clear yet.

5 The linear model and system inversion

The most popular model of the observations for source separation in the context of CMB observations probably is the linear mixture.

In this model, all components are assumed to have an emission which can be decomposed as the product of a spatial template independent of the frequency of observation, and of a spectral emission law which does not depend on the pixel. The total emission at frequency ν , in pixel p , of a particular emission process j is written as

$$x_j(\nu, p) = a(\nu)s_j(p)$$

or alternatively, in spherical harmonics space,

$$x_j(\nu, \ell m) = a(\nu)s_j(\ell m)$$

Forgetting for the moment annoying details concerning the response of the instrument (beams, frequency bands, etc...) the observation with a detector is then:

$$y_i(p) = \sum_j x_j(\nu_i, p) + n_i(p)$$

where $n_i(p)$ is the contribution of noise for detector i . For a set of detectors, this can be recast in a matrix–vector form as

$$\mathbf{y}(p) = \mathbf{A}\mathbf{s}(p) + \mathbf{n}(p) \quad (23)$$

Here, $\mathbf{y}(p)$ represent the set of maps observed with all detectors, and $\mathbf{s}(p)$ are the unobserved components (one template map per astrophysical component). The mixing matrix \mathbf{A} which does not depend on the pixel for a simple linear mixture, has one column per astrophysical component, and one line per detector.

If the observations are given in CMB temperature for all detectors, and if the detectors are properly calibrated, each element of the column of the mixing matrix corresponding to CMB is equal to 1.

The problem of component separation consists in inverting the linear system of equation 23. Here we first concentrate on linear inversion, which consists in finding the “best” possible matrix \mathbf{W} (such that $\hat{\mathbf{s}} = \mathbf{W}\mathbf{y}$ is “as good an estimator of \mathbf{s} as possible”).

Covariances and multivariate power spectra

In the following, a lot of use will be made of second order statistics of various sorts of data. In general, for a collection of maps $\mathbf{x}(p) = \{x_i(p)\}$, the covariance will be noted as $\mathbf{R}_{\mathbf{x}}(p, p')$, the elements of which are:

$$R_{ij}(p, p') = \text{cov}(x_i(p), x_j(p'))$$

Alternatively, in harmonic space, we denote as $\mathbf{R}_\mathbf{x}(\ell)$ the multivariate power spectrum of \mathbf{x} , i.e. the collection of matrices

$$\mathbf{R}_\mathbf{x}(\ell) = \langle \mathbf{x}(\ell, m) \mathbf{x}^\dagger(\ell, m) \rangle$$

where the brackets $\langle \cdot \rangle$ denote ensemble average, and the dagger † denotes the transpose of the complex conjugate. Such a power spectrum is well defined only for stationary/isotropic random fields on the sphere for which $\langle \mathbf{x}(\ell, m) \mathbf{x}^\dagger(\ell, m) \rangle$ does not depend on m .

5.1 Simple inversion

If \mathbf{A} is square and non singular, in absence of any additional information, then the inversion is obtained by

$$\mathbf{W} = \mathbf{A}^{-1} \quad (24)$$

and we have

$$\hat{\mathbf{s}} = \mathbf{A}^{-1} \mathbf{y} = \mathbf{s} + \mathbf{A}^{-1} \mathbf{n} \quad (25)$$

Note that because of the remaining noise term, this inversion is not always the best solution in terms of residual error, in particular in the poor signal to noise regimes. For instance, if we have two measurements of a mixture of CMB + thermal dust in a clean region of the sky (low foregrounds), one of which, at 150 GHz, is relatively clean, and the other, at 350 GHz, quite poor because of high level noise, then it may be better to use the 150 GHz as the CMB template (even with some dust contamination), rather than to invert the system, subtracting little dust and adding a large amount of noise.

In terms of residual foreground contamination however (if the criterion is to reject astrophysical signals, whatever the price to pay in terms of noise), the only solution here is matrix inversion. The solution is unbiased, but may be noisy.

Note that an ILC method would produce a different solution, possibly slightly biased (as discussed in 2.6), but possibly better in terms of signal to noise ratio of the end product.

This solution can be applied if the full matrix \mathbf{A} is known (not only the column of the component of interest, i.e. the CMB), without further prior knowledge of the data.

5.2 Inversion of a redundant system using the pseudo inverse

If there are more observations than components, but nothing is known about noise and signal levels, then the inversion is obtained by

$$\mathbf{W} = [\mathbf{A}^\dagger \mathbf{A}]^{-1} \mathbf{A}^\dagger \quad (26)$$

and we have

$$\hat{\mathbf{s}} = [\mathbf{A}^\dagger \mathbf{A}]^{-1} \mathbf{A}^\dagger \mathbf{y} = \mathbf{s} + [\mathbf{A}^\dagger \mathbf{A}]^{-1} \mathbf{A}^\dagger \mathbf{n} \quad (27)$$

Again, this estimator is unbiased, but may contain a large amount of noise and may not be optimal in terms of signal to noise ratio. All the comments made in the previous paragraph hold as well for this solution.

Note that there is no noise-weighting here, so that one single very bad channel may contaminate significantly all the data after inversion. It is therefore not a very good idea to apply this estimator with no further thoughts.

Note that, again, this solution can be implemented without any further knowledge about signal and noise – only the entries of the mixing matrix for all components are needed.

5.3 A noise-weighted scheme: the Generalised Least-Square solution

Let us now assume that we know something additional about the noise, namely, its second order statistics. These are described by noise correlation matrices in real space, or alternatively by noise power spectra in Fourier (for small maps) or in harmonic (for all-sky maps) space.

We denote as \mathbf{R}_n the noise correlation matrix and assume, for the time being, that the noise for each detector i is a realization of a random gaussian field, the generalised (or global) least square (GLS) solution of the system of equation 23 is:

$$\mathbf{W} = [\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1} \quad (28)$$

and we have

$$\hat{\mathbf{s}} = [\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{y} = \mathbf{s} + [\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{n} \quad (29)$$

Again, the solution is unbiased. Although there remains a noise contribution, this is the solution yielding the minimum variance error map for a deterministic signal (in contrast with the Wiener solution below, which optimises the variance of the error when the signal is stochastic, i.e. assumed to be a random field). It is also the best linear solution in the limit of large signal to noise ratio.

This solution is also theoretically better than the ILC when the model holds, but the price to pay is the need for more prior knowledge about the data (knowledge of the mixing matrix and of noise covariance matrices or power spectra). If that knowledge is insufficient, one has to design methods to get it from the data itself. Such “model learning” methods will be discussed in section 7.

5.4 The Wiener solution

The Wiener filter [Wiener (1949)] has originally been designed to filter time series in order to suppress noise, but has been extended to a large variety

of applications since then. Wiener's solution requires additional information regarding the spectral content of the original signal and the noise. Wiener filters are characterized by the following:

- Both the noise and the signal are considered as stochastic processes with known spectral statistics (or correlation properties) – contrarily to the GLS method which considers the noise only to be stochastic, the signal being deterministic,
- The optimization criterion is the minimum least square error,
- The solution is linear.

In signal processing, a data stream $y(t) = s(t) + n(t)$ assumed to be a noisy measurement of a signal s can be filtered for denoising as follows: in Fourier space, each mode $y(f)$ of the data stream is weighted by a coefficient

$$W(f) = \frac{S(f)}{S(f) + N(f)}$$

where $S(f) = \langle |s(f)|^2 \rangle$ and $N(f) = \langle |n(f)|^2 \rangle$ are ensemble averages of the square moduli of the Fourier coefficients of the stochastic processes s and n .

In the limit of very small noise level $N(f) \ll S(f)$, the Wiener filter value is $W(f) = 1$, and the filter does not change the data. In the limit of very poor signal to noise $S(f) \ll N(f)$, the filter suppresses the data completely, because that mode adds noise to the total data stream, and nothing else.

It can be shown straightforwardly that the Wiener filter minimizes the variance of the error of the signal estimator $\hat{s}(f) = W(f)y(f)$ (so that $\langle \int_f |\hat{s}(f) - s(f)|^2 df \rangle$ is minimal).

The Wiener solution can be adapted for solving our component separation problem, provided the mixing matrix \mathbf{A} and the second order statistics of the components and of the noise are known [Tegmark & Efstathiou (1996), Bouchet & Gispert (1999)] as:

$$\mathbf{W}^{(1)} = [\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} + \mathbf{R}_s^{-1}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1} \quad (30)$$

where \mathbf{R}_s is the correlation matrix of the sources (or power spectra of the sources, in the Fourier or harmonic space), and \mathbf{R}_n the correlation matrix of the noise. The superscript (1) is used to distinguish two forms of the Wiener filter (the second is given later in this section).

An interesting aspect of the Wiener filter is that:

$$\begin{aligned} \hat{\mathbf{s}} &= [\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} + \mathbf{R}_s^{-1}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{y} \\ &= [\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} + \mathbf{R}_s^{-1}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} \mathbf{s} \\ &\quad + [\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} + \mathbf{R}_s^{-1}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{n} \end{aligned} \quad (31)$$

The matrix in front of \mathbf{s} is not the identity, and thus the Wiener filter does not give an unbiased estimate of the signals of interest. Diagonal terms can be

different from unity. In addition, non-diagonal terms may be non-zero, which means that the Wiener filter allows some residual foregrounds to be present in the final CMB map – the objective being to minimise the variance of the residuals, irrespective of whether these residuals originate from instrumental noise or from astrophysical foregrounds.

As noted in [Tegmark & Efstathiou (1996)], the Wiener solution can be “debiased” by multiplying the Wiener matrix by a diagonal matrix removing the impact of the filtering. The authors argue that for the CMB this debiasing is desirable for subsequent power spectrum estimation on the reconstructed CMB map. Each mode of a given component is divided by the diagonal element of the Wiener matrix for that component and that mode. This, however, destroys the minimal variance property of the Wiener solution, and can increase the noise very considerably. There is an incompatibility between the objective of obtaining a minimum variance map, and the objective of obtaining an unbiased map which can be used directly to measure the power spectrum of the CMB. There is no unique method for both.

Before moving on, it is interesting to check that the matrix form of the Wiener filter given here reduces to the usual form when there is one signal only and when the matrix \mathbf{A} reduces to a scalar equal to unity. In that case, the Wiener matrix \mathbf{W} of equation 30 reduces to

$$W(f) = [1/S(f) + 1/N(f)]^{-1}/N(f) = S(f)/[N(f) + S(f)]$$

where S and N are the signal and noise power spectra, and we recover the classical Wiener formula.

Two forms of the Wiener Filter

In the literature, another form can be found for the Wiener filter matrix:

$$\mathbf{W}^{(2)} = \mathbf{R}_s \mathbf{A}^\dagger [\mathbf{R}_n + \mathbf{A} \mathbf{R}_s \mathbf{A}^\dagger]^{-1} \quad (32)$$

It can be shown straightforwardly that if the matrices

$$\mathbf{M}_1 = [\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} + \mathbf{R}_s^{-1}]$$

and

$$\mathbf{M}_2 = [\mathbf{R}_n + \mathbf{A} \mathbf{R}_s \mathbf{A}^\dagger]$$

are regular, then the forms of equation 30 and 32 are equivalent (simply multiply both forms by \mathbf{M}_1 on the left and \mathbf{M}_2 on the right, and expand).

It may seem that the form of equation 32 is more convenient, as it requires only one matrix inversion instead of three. Each form, however, presents specific advantages or drawbacks, which appear clearly in the high signal to noise ratio (SNR) limit, and if power spectra of all signals are not known.

The high SNR limit

The two above forms of the Wiener filter are not equivalent in the high SNR limit. In this regime, equation 30 yields in the limit

$$\mathbf{W}_{\text{limit}}^{(1)} = [\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1}$$

which is the GLS solution of equation 28, and depends only on the noise covariance matrix, whereas equation 32 tends to

$$\mathbf{W}_{\text{limit}}^{(2)} = \mathbf{R}_s \mathbf{A}^\dagger [\mathbf{A} \mathbf{R}_s \mathbf{A}^\dagger]^{-1}$$

which depends only on the covariance of the signal. Therefore, some care should be taken when applying the Wiener filter in the high SNR ratio regime, when numerical roundup errors may cause problems.

Note that if $[\mathbf{A}^\dagger \mathbf{A}]$ is regular, then

$$\begin{aligned} \mathbf{W}_{\text{limit}}^{(2)} &= \mathbf{R}_s \mathbf{A}^\dagger [\mathbf{A} \mathbf{R}_s \mathbf{A}^\dagger]^{-1} \\ &= [\mathbf{A}^\dagger \mathbf{A}]^{-1} [\mathbf{A}^\dagger \mathbf{A}] \mathbf{R}_s \mathbf{A}^\dagger [\mathbf{A} \mathbf{R}_s \mathbf{A}^\dagger]^{-1} \end{aligned} \quad (33)$$

$$= [\mathbf{A}^\dagger \mathbf{A}]^{-1} \mathbf{A}^\dagger \quad (34)$$

and the limit is simply the pseudo inverse of matrix \mathbf{A} , without any noise weighting. Of course, when there is no noise at all, $\mathbf{W}^{(1)}$ can not be implemented at all, and the Wiener solution is pointless anyway.

What if some covariances are not known?

It is interesting to note that even if the covariance matrix (or equivalently multivariate power spectrum) \mathbf{R}_s of all sources is not known, it is still possible to implement an approximate Wiener solution if the maps of observations are large enough to allow a good estimate of the covariance matrix of the observations.

If $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}$ and if the noise and the components are independent, the covariance \mathbf{R}_y of the observations is of the form:

$$\mathbf{R}_y = \mathbf{R}_n + \mathbf{A} \mathbf{R}_s \mathbf{A}^\dagger$$

Therefore, form 2 of the Wiener filter can be recast as:

$$\mathbf{W}^{(2)} = \mathbf{R}_s \mathbf{A}^\dagger [\mathbf{R}_y]^{-1} \quad (35)$$

If all components are decorrelated, the matrix \mathbf{R}_s is diagonal. For the implementation of a Wiener solution for one single component (e.g. CMB), only the diagonal element corresponding to the CMB (i.e. the power spectrum C_ℓ of the CMB) is needed, in addition to the multivariate power spectrum of the observations \mathbf{R}_y . The latter can be estimated directly using the observations.

5.5 Comment on the various linear inversion solutions

The above four linear solutions to the inversion of the linear system of equation 23 have been presented by order of increasing generality, increasing complexity, and increasing necessary prior knowledge. The various solutions are summarised in table 1. Three comments are necessary.

Firstly, we note that the Wiener solutions require the prior knowledge of the covariance matrices (or equivalently power spectra) of both the noise and the signal. For CMB studies, however, the measurement of the power spectrum of the CMB field is precisely the objective of the observations. Then, the question of whether the choice of the prior on the CMB power spectrum biases the final result or not is certainly of much relevance. For instance, the prior assumption that the power spectrum of the CMB is small in some ℓ range will result in filtering the corresponding modes, and the visual impression of the recovered CMB will be that indeed there is little power at the corresponding scales. For power spectrum estimation on the maps, however, this effect can be (and should be) corrected, which is always possible as the effective filter induced by the Wiener solution is known (for an implementation in harmonic space, it is equal for each mode ℓm , for each component, to the corresponding term of the diagonal of $\mathbf{W}_{\ell m} \mathbf{A}$). In section 8, a solution will be proposed for estimating first on the data themselves all the relevant statistical information (covariance matrices and frequency scalings), and then using this information for recovering maps of the different components.

Secondly, we should emphasise that the choice of a linear solution should be made with a particular objective in mind. If the objective is to get the best possible map in terms of least square error, then the Wiener solution is the best solution if the components are Gaussian. The debiased Wiener is not really adapted to any objective in particular. The GLS solution is the best solution if the objective is an unbiased reconstruction with no filtering and no contamination. In practice, it should be noted that small uncertainties on \mathbf{A} result in errors (biases and contamination) even for the GLS solution.

As a third point, we note that it can be shown straightforwardly that for Gaussian sources and noise, the Wiener solution maximises the posterior probability $P(\mathbf{s}|\mathbf{y})$ of the recovered sources given the data. From Bayes theorem, the posterior probability is the product of the likelihood $P(\mathbf{y}|\mathbf{s})$ and the prior $P(\mathbf{s})$, normalised by the evidence $P(\mathbf{y})$. The normalising factor does not depend on \mathbf{s} . We can write:

$$\begin{aligned} P(\mathbf{s}|\mathbf{y}) &\propto \exp [-(\mathbf{y} - \mathbf{A}\mathbf{s})^\dagger \mathbf{R}_n^{-1} (\mathbf{y} - \mathbf{A}\mathbf{s})] \exp [-\mathbf{s}^\dagger \mathbf{R}_s^{-1} \mathbf{s}] \\ \log(P(\mathbf{s}|\mathbf{y})) &= -[(\mathbf{y} - \mathbf{A}\mathbf{s})^\dagger \mathbf{R}_n^{-1} (\mathbf{y} - \mathbf{A}\mathbf{s})] - [\mathbf{s}^\dagger \mathbf{R}_s^{-1} \mathbf{s}] + \text{const.} \end{aligned} \quad (36)$$

where $\exp [-\mathbf{s}^\dagger \mathbf{R}_s^{-1} \mathbf{s}]$ is the Gaussian prior for \mathbf{s} . The requirement that

$$\frac{\partial}{\partial \mathbf{s}} \log(P(\mathbf{s}|\mathbf{y})) = 0$$

implies

$$\begin{aligned}
-\mathbf{A}^\dagger \mathbf{R}_n^{-1} (\mathbf{y} - \mathbf{A} \mathbf{s}) + \mathbf{R}_s^{-1} \mathbf{s} &= 0 \\
\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} \mathbf{s} + \mathbf{R}_s^{-1} \mathbf{s} &= \mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{y} \\
[\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} + \mathbf{R}_s^{-1}] \mathbf{s} &= \mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{y}
\end{aligned} \tag{37}$$

and thus we get the solution $\mathbf{W}^{(1)}$. In section 6, we will discuss the case where the Gaussian prior is replaced by an entropic prior, yielding yet another solution for \mathbf{s} .

5.6 Pixels, harmonic space, or wavelets?

The simple inversion of \mathbf{A} using the inverse or pseudo-inverse can be implemented equivalently with any representations of the maps, in pixel domain, harmonic space, or on any decomposition of the observations on a set of functions as, e.g., wavelet decompositions [Moudden et al. (2005)]. The result in terms of separation is independent of this choice, as far as the representation arises from a linear transformation.

If all sources and signals are Gaussian random fields, the same is true for GLS or Wiener inversions, provided all the second order statistics are properly described by the covariance matrices \mathbf{R}_n and \mathbf{R}_s .

These covariance matrices, in pixel space, take the form of the set of covariances:

$$\mathbf{R}_n = \{R_{n_i n_j}(p_i, p_j)\}$$

where

$$R_{n_i n_j}(p_i, p_j) = \langle n_i(p_i) n_j(p_j) \rangle$$

Similarly, in harmonic space, we have:

$$\mathbf{R}_n = \{R_{n_i n_j}(\ell_i, m_i, \ell_j, m_j)\}$$

where

$$R_{n_i n_j}(\ell_i, m_i, \ell_j, m_j) = \langle n_i(\ell_i, m_i) n_j(\ell_j, m_j) \rangle$$

If the number of pixels is large, if we deal with several sources and many channels at the same time (tens today, thousands in a few years), the implementation of the GLS or Wiener solution may be quite demanding in terms of computing. For this reason, it is desirable to implement the solution in the space where matrices are the most easy to invert.

For stationary Gaussian random fields, harmonic space implementations are much easier than direct space implementations, because the covariance between distinct modes vanish, so that

$$R_{n_i n_j}(\ell_i, m_i, \ell_j, m_j) = \langle n_i(\ell_i, m_i) n_j(\ell_i, m_i) \rangle \delta_{\ell_i \ell_j} \delta_{m_i m_j}$$

The full covariance matrix consists in a set of independent covariance matrices (one for each mode), each of which is a small matrix of size $N_{\text{channels}} \times N_{\text{channels}}$ for \mathbf{R}_n , and of size $N_{\text{sources}} \times N_{\text{sources}}$ for \mathbf{R}_s .

Solution	$\mathbf{W} =$	Required prior knowledge	Comments
Inverse	\mathbf{A}^{-1}	\mathbf{A}	When there are as many channels of observation as components. Unbiased, contamination free.
Pseudo-inverse	$[\mathbf{A}^\dagger \mathbf{A}]^{-1} \mathbf{A}^\dagger$	\mathbf{A}	When there are more channels of observation than components. Unbiased, contamination free.
GLS	$[\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1}$	\mathbf{A} and \mathbf{R}_n	Minimises the variance of the error for deterministic signals. Unbiased, contamination free.
Wiener 1	$[\mathbf{A}^\dagger \mathbf{R}_n^{-1} \mathbf{A} + \mathbf{R}_s^{-1}]^{-1} \mathbf{A}^\dagger \mathbf{R}_n^{-1}$	\mathbf{A} , \mathbf{R}_n and \mathbf{R}_s	Minimises the variance of the error for stochastic signals. Biased, not free of contamination. Tends to the GLS solution in the limit of high SNR.
Wiener 2	$\mathbf{R}_s \mathbf{A}^\dagger [\mathbf{R}_n + \mathbf{A} \mathbf{R}_s \mathbf{A}^\dagger]^{-1}$	\mathbf{A} , \mathbf{R}_n and \mathbf{R}_s	Equivalent to Wiener 1. Tends to the pseudo inverse in the limit of high SNR.
Debiased Wiener	$\Lambda \mathbf{R}_s \mathbf{A}^\dagger [\mathbf{R}_n + \mathbf{A} \mathbf{R}_s \mathbf{A}^\dagger]^{-1}$	\mathbf{A} , \mathbf{R}_n and \mathbf{R}_s	The diagonal matrix Λ , inverse of the diagonal of $\mathbf{W}\mathbf{A}$ where \mathbf{W} is the Wiener solution, removes for each mode the filtering effect of the Wiener filter. Unbiased, but not contamination free.

Table 1. Summary of linear solutions to component separation when the mixing matrix \mathbf{A} is known.

5.7 Annoying details

Under the assumption that the response of each detector i in the instrument can itself be decomposed in the product of a spectral response $h_i(\nu)$ and a frequency independent symmetrical beam B_i , the contribution of component j to the observation obtained with detector i is:

$$y_{ij}(\ell m) = \left[\int_{\nu} h_i(\nu) a(\nu) d\nu \right] B_{i,\ell} s_j(\ell m)$$

where $B_{i,\ell}$ are the coefficients of the expansion of the symmetric beam of detector i on Legendre polynomials.

The mixing matrix of this new linear model is seen to include a band integration, assumed to first order to be independent of ℓ , a the effect of a beam, which depends on ℓ . Both can be taken into account in a linear inversion, if known a priori.

6 The Maximum Entropy Method

The Wiener filter provides the best (in terms of minimum-variance, or maximum likelihood) estimate of the component maps if two main assumptions hold. Firstly, the observations should be a linear mixture of distinct emissions. Secondly, the components and the noise should be (possibly correlated) Gaussian stationary random processes.

Unfortunately, the sky is known to be neither Gaussian, nor stationary, with the possible exception of the CMB itself. Is this critical?

The Maximum Entropy Method (MEM) of component separation is a method which inverts the same linear system of component mixtures, but assumes non-gaussian probability distributions [Hobson et al. (1998)].

6.1 Maximum Entropy

The concept of entropy in information theory has been introduced by Shannon in 1948 [Shannon (1948)]. The entropy of a discrete random variable X on a finite set of possible values $\{x_i\}$ with probability distribution function $p(x_i) = p(X = x_i)$, is defined as:

$$H(X) = - \sum_{i=1}^N p(x_i) \log p(x_i) \quad (38)$$

The principle of maximum entropy is based of the idea that whenever there is some choice to be made about the distribution function of the random variable X , one should choose the least informative option possible. Entropy measures the amount of information in a probability distribution, and entropy maximisation is a way of achieving this.

For instance, in the absence of any prior information, the probability distribution which maximises the entropy of equation 38 is a distribution with uniform probability, $p(x_i) = 1/N$, i.e. the least informative choice of a probability distribution on the finite set $\{x_i\}$, where all outcomes are equally likely. This is the most natural choice if nothing more is said about the probability distribution.

In the opposite, a most informative choice would be a probability which gives a certain result, (for instance always $X = x_1$). This is a probability distribution which minimizes entropy.

In the continuous case where X can achieve any real value x with probability density $p(x)$, entropy can be defined as:

$$H(X) = - \int_{-\infty}^{\infty} p(x) \log p(x) dx \quad (39)$$

Of course, maximum entropy becomes really useful when there is also additional information available. In this case, the entropy must be maximized within the constraints given by additional information.

For instance, the maximum entropy distribution of a real random variable of mean μ and variance σ^2 is the normal (Gaussian) distribution :

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$$

For this reason, in absence of additional information about the probability distribution of a random variable of known mean and variance, it is quite natural, according to the Maximum Entropy principle, to assume a Gaussian distribution – which maximises the entropy, and hence corresponds to the least informative choice possible.

An other useful example is the maximum entropy distribution of a real positive random variable of mean μ , which is the exponential distribution :

$$p(x) = \frac{1}{\mu} \exp(-x/\mu)$$

6.2 Relative entropy

In fact, the differential entropy of equation 39 has an unpleasant property. It is not invariant under coordinate transformations (on spaces with more than one dimension).

The definition of *relative entropy* (or Kullback-Leibler divergence) between two distributions solves the issue. It can be interpreted as a measure of the amount of additional information one gets from knowing the actual (true) probability distribution $p(x)$, instead of an imperfect model $m(x)$, and is given by:

$$D_{\text{KL}}(p||m) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{m(x)} dx \quad (40)$$

Later in this paper (in section 8), we will make use of the Kullback-Leibler divergence for measuring the “mismatch” between two positive matrices \mathbf{R}_1 and \mathbf{R}_2 . It will actually correspond to the KL divergence between two Gaussian distributions with covariance matrices \mathbf{R}_1 and \mathbf{R}_2 .

The relative entropy is invariant under coordinate transformations (because both the ratio $p(x)/m(x)$ and $p(x)dx$ are invariant under coordinate transformations).

6.3 Component separation with the MEM

In principle, replacing the Gaussian prior by some other prior is perfectly legitimate. In practice, the choice of such a prior is not obvious, as the full statistical description of a complex astrophysical component is difficult to apprehend..

Following the maximum entropy principle, one may decide to use as a prior the distribution which maximises the entropy given a set of constraints. If the constraints are the value of the mean, and the variance, then the maximum entropy prior is the Gaussian prior.

Hobson and collaborators, in their MEM paper [Hobson et al. (1998)], argue that based on the maximum entropy principle, an appropriate prior for astrophysical components s is

$$p(s) = \exp[-\alpha S_c(\mathbf{s}, \mathbf{m}_u, \mathbf{m}_v)] \quad (41)$$

with

$$S_c(\mathbf{s}, \mathbf{m}_u, \mathbf{m}_v) = \sum_{j=1}^L \left\{ \psi_j - m_{uj} - m_{vj} - s_j \ln \left[\frac{\psi_j + s_j}{2m_{uj}} \right] \right\}$$

where $\psi_j = [s_j^2 + 4m_{uj}m_{vj}]^{1/2}$, and where \mathbf{m}_u and \mathbf{m}_v are models of two positive additive distributions (which are not clearly specified) used to represent the astrophysical components.

A derivation for this is given in [Hobson & Lasenby (1998)], but the connection to entropy is not direct. In particular, the definition of entropy does not require the *values* of the random variables to be positive, but their *probability densities*, which makes the discussion unconvincing.

Pragmatically, the choice for the prior of equation 41 seems to be validated a posteriori by the performance of the separation, which is not worse (and actually better for some of the components) than that obtained with the Wiener filter. It is not likely to be optimal, however, because the non-stationarity of components implies correlations in the harmonic domain, which are not fully taken into account in the MEM implementation.

The maximisation of the posterior probability (and hence of the product of the likelihood and the prior), is done with a dedicated fast maximisation

algorithm. We refer the reader to the relevant papers for additional details [Hobson & Lasenby (1998), Stolyarov et al. (2002)].

This method has been applied to the separation of components in the COBE data [Barreiro et al. (2004)].

6.4 Comments about the MEM

Although entropy has a clear meaning in terms of information content in the discrete case (e.g. it defines the minimum number of bits necessary to represent a sequence), there is no such interpretation in the continuous case. Entropy maximisation, understood as minimising the amount of arbitrary information in the assumed distribution, hence, is not very clearly founded for continuous images.

The “principle” of maximum entropy, as the name indicates, is not a theorem, but a reasonable recipe which seems to work in practice. In the context of the CMB, there is no guarantee that it is optimal, among all non-linear solutions of the mixing system. MEM outperforms the Wiener filter solution for some components in particular because the entropic prior of Hobson and Lasenby allows heavier tails than the Gaussian prior. Other priors however, based on a physical model of the emissions, might well perform even better in some cases. This question remains as an open problem in the field.

7 ICA and Blind source separation

7.1 About blind separation

The term “blind separation” refers to a fascinating possibility: if the components of a linear mixture are *statistically independent*, they can be recovered even if the mixing matrix \mathbf{A} is unknown *a priori*. In essence, this is possible because statistical independence is, at the same time, a strong mathematical property and, quite often, a physically plausible one.

There is an obvious and strong motivation for attempting blind component separation: allowing underlying components to be recovered blindly makes it possible to analyze multi-detector data with limited, imperfect, or even outright missing knowledge about the emission laws of the components. Even better, one can process data without knowing in advance which components might be “out there”. Hence, the blind approach is particularly well suited for exploratory data analysis.

In the last fifteen years, blind component separation has been a very active area of research in the signal processing community where it goes by the names of “blind source separation” (BSS) and “independent component analysis” (ICA). This section outlines the principles underlying some of the best known approaches to blind source separation. There is not a single best approach because there is not a unique way in which to express statistical independence on the basis of a finite number of samples.

7.2 Statistical independence

This section explains why blind component separation is possible in the first place. For the sake of exposition, the main ideas are discussed in the simplest case: there is no observation noise and there are as many “channels” as underlying components. Thus the model reduces to

$$\mathbf{y}(p) = \mathbf{A}\mathbf{s}(p)$$

where \mathbf{A} is an $n \times n$ matrix and we are looking for an $n \times n$ matrix “separating matrix” \mathbf{W} . Of course, if the mixing matrix \mathbf{A} is known, there is little mystery about separation: one should take $\mathbf{W} = \mathbf{A}^{-1}$ and be done with it.

If nothing is known about \mathbf{A} but the components are known (or assumed) to be statistically independent, the idea is to determine \mathbf{W} in such a way that the entries of vector $\mathbf{W}\mathbf{y}$ are independent (or as independent as possible). In other words, the hope is that by restoring independence, one would restore the components themselves. Amazingly enough, this line of attack works. Even better, under various circumstances, it can be shown to correspond to maximum likelihood estimation and there is therefore some statistical optimality to it. . . provided the hypothesis of statistical independence is expressed vehemently enough.

Note however that no matter the amount of statistical ingenuity thrown at blind component separation, there is no hope to recover completely the mixing matrix (or equivalently: the components). This is because a scalar factor can always be exchanged between each entry of \mathbf{s} and the corresponding column of \mathbf{A} without changing what the model predicts (*i.e.* the value of the product $\mathbf{A}\mathbf{s}$) and without destroying the (hypothetical) independence between the entries of \mathbf{s} . The same is true of a renumbering of the columns of \mathbf{A} and of the entries of \mathbf{s} . In other words, blind recovery is possible only up to rescaling and permutation of the components. In many applications, this will be “good enough”. If these indeterminacies have to be fixed, it can be done only by imposing additional constraints or resorting to side information.

For any possible choice \mathbf{W} of a candidate separating matrix, denote

$$\mathbf{x}(p) = \mathbf{W}\mathbf{y}(p)$$

the corresponding vector of candidate components. If $\mathbf{W} = \mathbf{A}^{-1}$ then the entries of \mathbf{x} are independent (since, in this case $\mathbf{x}(p) = \mathbf{s}(p)$). Under which circumstances would the *converse* true? Whenever the converse is true, it will be possible to recover the sources by looking for the linear transform \mathbf{W} which makes them independent. Hence, we have a blind separation principle: to separate components, make them independent.

7.3 Correlations

The main difficulty in blind source separation is to define a measure of independence. The problem is that the simple decorrelation condition⁴ between any two candidate components:

$$\frac{1}{P} \sum_{p=1}^P x_i(p)x_j(p) = 0 \quad \text{for } 1 \leq i \neq j \leq n. \quad (42)$$

does not cut it. This is in fact obvious from the fact that this decorrelation condition between x_i and x_j is *symmetric*. Hence decorrelation provides only $n(n-1)/2$ constraints while n^2 constraints are needed to determine \mathbf{W} uniquely. Therefore, more expressive forms of independence must be used. Two main avenues are possible: non-linear correlations and localized correlations, as described next.

Non-linear correlations

The “historical approach” to blind separation has been to determine a separating matrix \mathbf{W} in order to obtain “non-linear decorrelations” *i.e.*

$$\frac{1}{P} \sum_{p=1}^P \psi_i(x_i(p)) x_j(p) = 0 \quad \text{for } 1 \leq i \neq j \leq n \quad (43)$$

where functions $\psi_1, \dots, \psi_n : R \mapsto R$ are *non-linear* functions (more about choosing them below). By using non-linear functions, symmetry is broken and the required number of constraints is obtained, namely $n(n-1)$ (with n additional arbitrary constraints, needed for fixing the scale of each component).

Localized correlations

Another approach is to look for “localized decorrelation” in the sense of solving

$$\frac{1}{P} \sum_{p=1}^P \frac{x_i(p)}{\sigma_{ip}^2} x_j(p) = 0 \quad \text{for } 1 \leq i \neq j \leq n \quad (44)$$

where for each component i , a sequence $\{\sigma_{ip}\}_{p=1}^P$ of positive number must be defined (more about this soon). Again, blind identification is possible because symmetry is broken, provided no two sequences of σ ’s are proportional.

⁴ Here, as in the rest of this section, all signals are assumed to have zero mean.

Maximum likelihood

Why using the particular proposals (43) or (44) as extended decorrelation conditions rather than any other form, possibly more complicated? One reason is that reasonable algorithms exist for computing the \mathbf{W} such that $\mathbf{x} = \mathbf{W}\mathbf{y}$ is a solution of (43) or (44). Another, more important reason is that these two conditions actually characterize the maximum likelihood estimate of \mathbf{W} in simple and well understood models. Because of this, we can understand what the algorithm does and we have guidance for choosing the non-linear functions ψ_i in condition (43) or the varying variance profiles σ_{iq}^2 in condition (44) as stated next.

Non linear correlations. Assume that each component $\{s_i(p)\}$ is modeled as having all pixels independently and identically distributed according to some probability density p_i . In this model, the most likely value of \mathbf{A} given the observations has for inverse a matrix \mathbf{W} such that condition (43) holds with $\psi_i = -p'_i/p_i$. Hence, if the model is true (or approximately true), the non linear-function appearing in condition (43) should be taken as minus the derivative of the log-density of $s_i(p)$. For a Gaussian distribution p_i , the corresponding function ψ_i is linear: here, the necessary non-linearity of ψ_i corresponds to the non Gaussianity of the corresponding component.

Localized correlations. Alternatively, one may model each component $\{s_i(p)\}$ as having all pixels independently and *normally* distributed with zero-mean and “local” variance σ_{ip}^2 . Then, in this model, the likeliest value of \mathbf{A} given the observations has for inverse a matrix \mathbf{W} such that $\mathbf{x} = \mathbf{W}\mathbf{y}$ satisfies condition (44).

7.4 ICA in practice

For the simple noise-free setting under consideration (the noisy case is addressed in next section), the algorithmic solutions depend on the type of decorrelation one decides to use.

Non linear decorrelation

Two popular ICA algorithms based on non-linear decorrelation (hence exploiting non Gaussianity) are JADE [Cardoso (1999)] and FastICA [Hyvärinen (1999)]. In practice however, these algorithms do not exactly solve an equation in the form (43). Rather, for algorithmic efficiency, they try to solve it under the additional constraint that the components are uncorrelated *i.e.* that condition (42) is satisfied exactly. The underlying optimization engine is a joint diagonalization algorithm for JADE and a fixed point technique for FastICA.

Localized decorrelation

Efficient algorithms for solving the localized decorrelation conditions (44) are based on assuming some regularity in the variance profiles: the sequences

$\{\sigma_{ip}^2\}$ are approximated as being constant over small domains. Hence, the global set $[1, P]$ is partitioned into Q subsets $\mathcal{I}_1, \dots, \mathcal{I}_Q$, each containing a number P_q of points (so that $P = \sum_{q=1}^Q P_q$). In practice, these pixel subsets are (well chosen) spatial regions. With a slight abuse of notation, we write $\sigma_{ip}^2 = \sigma_{iq}^2$ if $p \in \mathcal{I}_q$. Then, a small amount of maths turns the decorrelation conditions (44) into

$$\left[\sum_{q=1}^Q P_q \Sigma_q^{-1} \mathbf{W} \hat{\mathbf{R}}_q \mathbf{W}^\dagger \right]_{ij} = 0 \quad \text{for } 1 \leq i \neq j \leq n \quad (45)$$

where $\hat{\mathbf{R}}_q$ is a localized covariance matrix

$$\hat{\mathbf{R}}_q = \frac{1}{P_q} \sum_{p \in \mathcal{I}_q} \mathbf{y}(p) \mathbf{y}(p)^\dagger \quad \text{and} \quad \Sigma_q = \text{diag}(\sigma_{1q}^2, \dots, \sigma_{nq}^2). \quad (46)$$

An important point here is that by assuming piecewise constant variance profiles, the localized decorrelation condition can be expressed entirely in terms of the localized covariance matrices $\hat{\mathbf{R}}_q$. Hence the localized covariance matrices appear as *sufficient statistics* in this model. Even better, the likelihood of \mathbf{A} can be understood as a mismatch between these statistics and their predicted form, namely $\mathbf{R}_q = \mathbf{A} \Sigma_q \mathbf{A}^\dagger$. Specifically, in this model the probability $p(\mathbf{y}(1), \dots, \mathbf{y}(P) | \mathbf{A}, \Sigma)$ of the data given \mathbf{A} and the set $\Sigma = \{\Sigma_1, \dots, \Sigma_Q\}$ of covariance matrices is given by

$$\log p(\mathbf{y}(1), \dots, \mathbf{y}(P) | \mathbf{A}, \Sigma) = -\phi(\mathbf{A}, \Sigma) + \text{cst}$$

where function ϕ is defined as

$$\phi(\mathbf{A}, \Sigma) = \sum_q P_q K(\hat{\mathbf{R}}_q, \mathbf{A} \Sigma_q \mathbf{A}^\dagger) \quad (47)$$

and where $K(\cdot, \cdot)$ is a measure of divergence between two matrices defined as

$$K(\mathbf{R}_1, \mathbf{R}_2) = \frac{1}{2} (\text{trace}(\mathbf{R}_1 \mathbf{R}_2^{-1}) - \log \det(\mathbf{R}_1 \mathbf{R}_2^{-1}) - n) \quad (48)$$

This shows that maximum likelihood estimation of \mathbf{A} amounts to the minimization of the weighted mismatch (47) between the set of localized covariance matrices $\hat{\mathbf{R}}_q$ (computed from the data) and their expected value $\mathbf{R}_q = \mathbf{A} \Sigma_q \mathbf{A}^\dagger$ (predicted by the model).

In the noise-free case considered here, it turns out that there is a simple and very efficient algorithm (due to D.T. Pham) for minimizing the spectral mismatch.

8 SMICA

We have developed a component separation technique dubbed SMICA for ‘spectral matching ICA’ which is based on the ideas sketched at previous section but improves on them in several ways.

In its simplest form, SMICA is based on *spectral* statistics, that is, on statistics which are localized not in space but in frequency. These statistics are binned auto- and cross-spectra of the channels. More specifically, for a given set of N_{chann} multi-channels maps $\{y_i(p)\}$, we form for each (ℓ, m) the $N_{\text{chann}} \times 1$ vector $\mathbf{y}(\ell, m)$ of their harmonic coefficients and define

$$\hat{\mathbf{R}}(\ell) = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{m=+\ell} \mathbf{y}(\ell, m) \mathbf{y}(\ell, m)^\dagger$$

These empirical spectral covariance matrices are then binned. In the simplest case, we define Q top-hat bins, with the q -th frequency bin contains all frequencies ℓ between ℓ_q^{\min} and ℓ_q^{\max} . We consider the binned spectra:

$$\hat{\mathbf{R}}_q = \frac{1}{P_q} \sum_{\ell=\ell_q^{\min}}^{\ell=\ell_q^{\max}} (2\ell + 1) \hat{\mathbf{R}}(\ell) \quad \text{where} \quad P_q = \sum_{\ell=\ell_q^{\min}}^{\ell=\ell_q^{\max}} (2\ell + 1).$$

Here P_q is the number of Fourier modes summed together in a single estimate $\hat{\mathbf{R}}_q$.

The mixture model $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}$ predicts that the empirical spectra \mathbf{R}_q have an expected value

$$\mathbf{R}_q = \langle \hat{\mathbf{R}}_q \rangle = \mathbf{A} \Sigma_q \mathbf{A}^\dagger + \mathbf{N}_q$$

where Σ_q are the binned spectral covariance matrix for the components in bin q and \mathbf{N}_q is the same for noise, assumed to be uncorrelated from the components. The unknown parameters can be collected in a big vector θ :

$$\theta = \{\mathbf{A}, \{\Sigma_q\}, \{\mathbf{N}_q\}\}$$

but in practice we will not fit such a large model. Many constraints can be imposed on θ . A typical choice is to assume that the components are uncorrelated between themselves and that the noise also is uncorrelated between channels. Such a choice would result in a smaller parameter set

$$\theta = \{\mathbf{A}, \{\text{diag} \Sigma_q\}, \{\text{diag} \mathbf{N}_q\}\}$$

but infinitely many other options are possible, both more stringent (like assuming that the noise in each channel is a smooth function of the bin index q) or less stringent (like assuming that some components may not be uncorrelated). In the following, we do not assume a specific parametrization of

the binned spectral covariance matrices. Rather, we denote where θ is some parameter set which uniquely determines the values of \mathbf{A} and each \mathbf{R}_q and \mathbf{N}_q :

$$\{\mathbf{R}_q\} = \{\mathbf{R}_q(\theta)\} = \{\mathbf{A}(\theta)\mathbf{\Sigma}_q(\theta)\mathbf{A}(\theta)^\dagger + \mathbf{N}_q(\theta)\}$$

SMICA determines the set θ of unknown parameters by fitting the empirical spectral covariance matrices to whichever structure is predicted by the model. Specifically, the unknown parameters are found by minimizing the “spectral mismatch”

$$\phi(\theta) = \sum_q P_q K(\hat{\mathbf{R}}_q, \mathbf{R}_q(\theta)) \quad (49)$$

averaged across bins. Some comments are in order regarding the matching criterion, the issue of non stationarity and practical implementation.

Matching criterion

The reason for choosing this particular form of mismatch between data and model is that minimizing (49) is identical to maximizing the likelihood of the data in a model where all components are 1) Gaussian 2) stationary and 3) have harmonic spectra which are constant over bins. Of course, these assumptions are not met in practice so one could choose a different criterion for matching $\hat{\mathbf{R}}_q$ to $\mathbf{R}_q(\theta)$ but we have little statistical guidance for picking up an alternate matching measure. Furthermore, the assumptions 1) and 2) are met by the CMB and 3) is approximately correct for narrow bins. In addition, the failure of stationarity can be alleviated by using localized statistics (see below).

Non stationarity and localization

The spectral approach to building a likelihood function has some benefits, in particular the fact that it is perfectly suited to describing the statistical properties of the CMB. Another beneficial side effect is that it makes it easy to deal with varying resolution from channel to channel as long as the beam can be considered to be symmetrical

However, going straight away to harmonic space seems unreasonable to deal with highly non stationary components such as the galactic components. This issue can be addressed to some extent by resorting to localized spectral statistics. It is a simple matter to use spatial window functions to partition the sky into spatial domains [Cardoso et al. (2005)]. Although not a perfect solution, it certainly allows to capture a good deal of the non stationary features of the galactic sky.

Implementation

The definition of the spectral matching criterion (49) encapsulates all of the statistical modeling but leaves open the separate and possibly tricky issue

of minimizing $\phi(\theta)$.⁵ Because the criterion is a likelihood in disguise, it is possible to use the EM algorithm for its minimization, with the components taken as latent variables. However, EM is often not fast enough and also is not able to deal with arbitrary parametrization of $\Sigma_q(\theta)$ and $\mathbf{N}_q(\theta)$. It has been found necessary to use general optimization techniques. A conjugate gradient algorithm can be implemented because a reasonably tractable expression for the gradient of the criterion is available as:

$$\frac{\partial \phi}{\partial \theta} = \sum_q P_q \text{trace} \left(\mathbf{R}_q(\theta)^{-1} (\mathbf{R}_q(\theta) - \hat{\mathbf{R}}_q) \mathbf{R}_q(\theta)^{-1} \frac{\partial \mathbf{R}_q(\theta)}{\partial \theta} \right)$$

However, in our context, the conjugate gradient algorithm also requires preconditioning. A preconditioner can be classically obtained as the inverse of the Fisher information matrix $\text{FIM}(\theta)$ which is taken as an approximation to the Hessian of $\phi(\theta)$:

$$\frac{\partial^2 \phi}{\partial \theta^2} \approx \text{FIM}(\theta) = \sum_q P_q \text{trace} \left(\mathbf{R}_q(\theta)^{-1} \frac{\partial \mathbf{R}_q(\theta)}{\partial \theta} \mathbf{R}_q(\theta)^{-1} \frac{\partial \mathbf{R}_q(\theta)}{\partial \theta} \right)$$

Mismatch control and error bars

A benefit of the SMICA approach is that it comes with a built-in measure of the quality of the model. Indeed, if we properly fit all the auto-cross spectra, then $\phi(\theta)$ should be ‘statistically small’. Visual control of the quality of the spectral matching is obtained by plotting $\phi_q = P_q K(\hat{\mathbf{R}}_q, \mathbf{R}_q(\hat{\theta}))$ versus q where $\hat{\theta}$ is the minimizer of $\phi(\theta)$. This quantity should be understood as a χ^2 . If the model holds (Gaussian stationary components and noise) and when all spectral parameters are freely estimated ϕ_q behaves approximately as a χ^2 with a number of degrees of freedom equal to $N_1 - N_2$ where $N_1 = N_{\text{chann}}(N_{\text{chann}} + 1)/2$ is the number of degrees of freedom for a sample covariance matrix of size N_{chann} and where $N_2 = N_{\text{comp}} + N_{\text{chann}}$ is the number of adjustable spectral parameters (the variances of each components and noise levels in a given frequency bin).

9 Other blind, semi-blind, or model learning methods

This paper would not be complete without a quick review of some of the recent work. We quote here a few papers which we think deserve reading for

⁵ We note in passing that some authors seem to make a confusion between the objective function (the criterion which has to be minimised, which derives from a statistical model) and the algorithm used for minimization. For instance, some authors use the terms “EM method”, or “MCMC method”, to design a method in which they use the EM algorithm, or Monte-Carlo Markov Chains. This is rather unfortunate, and contributes to a certain level of confusion.

further exploration of component separation issues and methods. Although unevenly mature, these methods provide complementary approaches, with advantages and drawbacks which deserve to be investigated.

9.1 FastICA

A blind component separation based on the FastICA method has been developed for CMB data reduction by Baccigalupi et al [Baccigalupi et al.(2000)], with an extension to the full sky by Maino et al [Maino et al.(2002)]. This blind approach uses, as “engine” for component separation, a measure of independence based on non-Gaussianity. Therefore, it is essentially equivalent to finding components which cancel non-linear correlations in the sense of equation 43.

For CMB applications, characterizing independence via non linear correlations of the form 43 has some limitations. Firstly, theory shows that this characterization allow for the separation of at most one Gaussian component [Cardoso (1998)]. The Gaussian component is somehow found “by default”, as the particular component which is orthogonal to (uncorrelated with) all other non Gaussian components. This is a concern for component separation performed with the CMB as the main target. Secondly, the non-linear decorrelation conditions do not take the noise into account. Even though this can be fixed in some ad hoc fashion, it is computationally demanding to do it in maximum likelihood sense. Finally, pixel space implementations cannot easily handle channel-dependent beams (unless explicit beam deconvolution is performed). If, to circumvent this problem, one considers harmonic space implementation, performance suffers from the fact that Fourier tend to be more Gaussian than the original, pixel-domain maps.

FastICA, however, can outperform other component separation methods for some applications. Spectral based methods (like SMICA) cannot blindly separate two components if their angular power spectra are proportional. FastICA does not suffer from this limitation and therefore has an edge for separating galactic components. If all galactic components have similar power spectra (say, proportional to ℓ^{-3}) then SMICA is expected to perform poorly without prior information.

Although both FastICA and SMICA are blind methods entering in the general class of “independent component analysis”, it should thus be stressed that they are conceptually very different. Performance, therefore, is expected to be very different also, and to depend on the actual properties of the data sets.

FastICA has been used on COBE and on WMAP data [Maino et al.(2003), Maino et al.(2007)].

9.2 Other recent developments

A “semi-blind” approach to component separation has been proposed by Tegmark and collaborators in a work where they model the foreground emissions using a number of physical parameters, which they estimate directly in the data sets [Tegmark et al.(2000)]. They estimate the impact of estimating these extra parameters in terms of accuracy loss on parameters of interest for CMB science. This paper was the first to address seriously the problem of component spectral indices varying over the sky.

Martinez-Gonzalez and collaborators have proposed a method for the extraction of the CMB specifically and for the estimation of its power spectrum [Martínez-González et al.(2003)]. The EM algorithm is the main tool of the implementation.

Eriksen and collaborators have developed a method based on a refined modeling of the astrophysical components, and fitting this model to the data to obtain estimates of foreground parameters [Eriksen et al. (2006)]. The fit of the parameters is made pixel by pixel at low-resolution using a MCMC technique for exploring the likelihood. After this first “model learning” step, the parameters obtained are used to estimate high resolution component maps.

Recently, Hansen and collaborators have proposed a CMB cleaning method based on a wavelet fit of component emissions obtained by differencing observations in different channels, and subtraction of the fit from observations made at frequencies where the CMB dominates [Hansen et al.(2006)].

Bonaldi and collaborators have recently published a paper for estimating parameters of emission of astrophysical components (emission laws, described by spectral indices). The statistics used are based on estimations of the correlations of the observations using a subset of points on the sphere [Bonaldi et al.(2006)].

An alternate way of performing component separation has been proposed by Bobin and collaborators, based on sparse representations of the various emissions [Bobin (2006)]. The basic principle of this method consists in decomposing the observations in a set of (redundant) dictionnaires chosen so that each component can be represented sparsely in one of the dictionnaires. Separation is achieved by minimizing the number of coefficients required to represent the data set.

A comparison of these different methods on a common data set, for investigating their strengths and weaknesses and evaluating their relative performance for various objectives would be an interesting work to improve the quality of component separation with the data set of upcoming space missions.

10 Conclusion and prospects

With improving data quality and increasingly demanding performance in component characterisation, component separation will play an important role in the analysis of CMB data sets in the next decade.

In this paper, we have reviewed the main issues for component separation, concentrating on diffuse components specifically.

Although substantial work has been performed, open questions remain. Polarisation, for instance, is one of the next major objectives of CMB science, for which much better sensitivities are required, and for which foreground emission is poorly known... Time varying sources, as the emission due to zodiacal light (modulated by the trajectory of the instrument in the ecliptic), as solar system objects in general, and as intrinsically time-varying radio sources, require specific methods tailored for their extraction.

The upcoming Planck data set, expected to become available to the Planck consortium in 2008, will provide a fantastic and challenging data set for extracting the emission from all astrophysical processes emitting in the millimeter range.

References

- [Aghanim et al. (1996)] Aghanim, N., Desert, F. X., Puget, J. L., & Gispert, R. 1996, *Ionization by early quasars and cosmic microwave background anisotropies*, Astronomy and Astrophysics, 311, 1
- [Audit & Simmons (1999)] Audit, E., & Simmons, J. F. L. 1999, *The kinematic Sunyaev-Zel'dovich effect and transverse cluster velocities*, MNRAS, 305, L27
- [Baccigalupi et al.(2000)] Baccigalupi, C., et al. 2000, *Neural networks and the separation of cosmic microwave background and astrophysical signals in sky maps*, MNRAS, 318, 769
- [Barreiro (2005)] Barreiro, R. B. 2006, *Techniques for compact source extraction on CMB maps*, to appear in "Data Analysis in Cosmology", Springer-Verlag Lecture Notes in Physics. Valencia, 6-10 September 2004 (ArXiv Astrophysics e-prints, arXiv:astro-ph/0512538)
- [Barreiro et al. (2004)] Barreiro, R. B., Hobson, M. P., Banday, A. J., Lasenby, A. N., Stolyarov, V., Vielva, P., & Górski, K. M. 2004, *Foreground separation using a flexible maximum-entropy algorithm: an application to COBE data*, MNRAS, 351, 515
- [Bennett et al. (2003)] Bennett, C. L., et al. 2003, *First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results*, ApJ Supplement Series, 148, 1
- [Bennett et al. (2003)] Bennett, C. L., et al. 2003, *First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: foreground emission*, ApJ Supplement Series, 148, 97
- [Benoît et al. (2004)] Benoît, A., et al. 2004, *First detection of polarization of the submillimetre diffuse galactic dust emission by Archeops*, Astronomy and Astrophysics, 424, 571

- [Birkinshaw (1999)] Birkinshaw, M. 1999, *The Sunyaev Zel'dovich effect* Physics Reports, 310, 97
- [Bobin (2006)] Bobin, J., Moudden, Y., Starck, J.-L. and Elad, M. 2006, *Morphological Diversity and Source Separation* IEEE Transaction on Signal Processing, in press.
- [Bonaldi et al.(2006)] Bonaldi, A., Bedini, L., Salerno, E., Baccigalupi, C., & de Zotti, G. 2006, *Estimating the spectral indices of correlated astrophysical foregrounds by a second-order statistical approach*, MNRAS, 373, 271
- [Bouchet & Gispert (1999)] Bouchet, F. R., & Gispert, R. 1999, *Foregrounds and CMB experiments I. Semi-analytical estimates of contamination*, New Astronomy, 4, 443
- [Cardoso (1998)] Cardoso, J.-F. 1998, *Blind signal separation: statistical principles*, Proceedings of the IEEE. Special issue on blind identification and estimation, vol. 9, no 10, pp. 2009-2025
- [Cardoso (1999)] Cardoso, J.-F. 1999, *High-order contrasts for independent component analysis*, Neural Computation, vol. 11, no 1, pp. 157-192
- [Cardoso et al. (2005)] Cardoso, J.-F., et al. 2005, *Statistiques direction-multipôle pour la séparation de composantes dans le fonds de rayonnement cosmologique*, Actes du GRETSI, Louvain-la-Neuve, Belgique
- [de Bernardis et al. (2000)] de Bernardis, P., et al. 2000, *A flat Universe from high-resolution maps of the cosmic microwave background radiation*, Nature, 404, 955
- [Delabrouille et al. (2003)] Delabrouille, J., Cardoso J.-F. and Patanchon, G. 2003, *Multi-detector multi-component spectral matching and application for CMB data analysis*, MNRAS, vol. 346, no 4, pp. 1089-1102
- [Dickinson et al. (2004)] Dickinson, C., et al. 2004, *High-sensitivity measurements of the cosmic microwave background power spectrum with the extended Very Small Array*, MNRAS, 353, 732
- [Dickinson et al. (2006)] Dickinson, C., Casassus, S., Pineda, J. L., Pearson, T. J., Readhead, A. C. S., & Davies, R. D. 2006, *An Upper Limit on Anomalous Dust Emission at 31 GHz in the Diffuse Cloud [LPH96] 201.663+1.643*, Astrophysical Journal Letters, 643, L111
- [Draine & Lazarian (1998)] Draine, B. T., & Lazarian, A. 1998, *Diffuse Galactic Emission from Spinning Dust Grains*, Astrophysical Journal Letters, 494, L19
- [Eriksen et al. (2004)] Eriksen, H. K., Banday, A. J., Górski, K. M., & Lilje, P. B. 2004, *On foreground removal from the WMAP data by an ILC method: limitations and implications*, ApJ, 612, 633
- [Eriksen et al. (2006)] Eriksen, H. K., et al. 2006, *Cosmic Microwave Background Component Separation by Parameter Estimation*, ApJ, 641, 665
- [Fernández-Cerezo et al. (2006)] Fernández-Cerezo, S., et al. 2006, *Observations of the cosmic microwave background and galactic foregrounds at 12-17GHz with the COSMOSOMAS experiment*, MNRAS, 370, 15
- [Gruzinov & Hu (1998)] Gruzinov, A., & Hu, W. 1998, *Secondary Cosmic Microwave Background Anisotropies in a Universe Reionized in Patches*, The Astrophysical Journal, 508, 435
- [Hansen et al.(2006)] Hansen, F. K., Banday, A. J., Eriksen, H. K., Górski, K. M., & Lilje, P. B. 2006, *Foreground Subtraction of Cosmic Microwave Background Maps Using WI-FIT (Wavelet-Based High-Resolution Fitting of Internal Templates)*, ApJ, 648, 784

- [Haslam et al. (1981)] Haslam, C. G. T., Klein, U., Salter, C. J., Stoffel, H., Wilson, W. E., Cleary, M. N., Cooke, D. J., & Thomasson, P. 1981, *A 408 MHz all-sky continuum survey. I - Observations at southern declinations and for the North Polar region*, *Astronomy and Astrophysics*, 100, 209
- [Hinshaw et al. (2006)] Hinshaw, G., et al. 2006, *Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Temperature Analysis*, ArXiv Astrophysics e-prints, arXiv:astro-ph/0603451
- [Hobson & Lasenby (1998)] Hobson, M. P., & Lasenby, A. N. 1998, *The entropic prior for distributions with positive and negative values*, *MNRAS*, 298, 905
- [Hobson et al. (1998)] Hobson, M. P., Jones, A. W., Lasenby, A. N., & Bouchet, F. R. 1998, *Foreground separation methods for satellite observations of the cosmic microwave background*, *MNRAS*, 300, 1
- [Hu & Dodelson (2002)] Hu, W., & Dodelson, S. 2002, *Cosmic Microwave Background Anisotropies*, *Annual Review of Astronomy and Astrophysics*, 40, 171
- [Hyvärinen (1999)] Hyvärinen, A. 1999, *Fast and Robust Fixed-Point Algorithms for Independent Component Analysis*, *IEEE Transactions on Neural Networks* 10(3):626-634
- [Jones et al. (2005)] Jones, W. C., et al. 2005, *A Measurement of the Angular Power Spectrum of the CMB Temperature Anisotropy from the 2003 Flight of Boomerang*, ArXiv Astrophysics e-prints, arXiv:astro-ph/0507494
- [Kogut et al. (1996)] Kogut, A., Banday, A. J., Bennett, C. L., Gorski, K. M., Hinshaw, G., & Reach, W. T. 1996, *High-Latitude Galactic Emission in the COBE Differential Microwave Radiometer 2 Year Sky Maps*, *The Astrophysical Journal*, 460, 1
- [Komatsu et al. (2003)] Komatsu, E., et al. 2003, *First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Tests of Gaussianity*, *ApJ Supplement Series*, 148, 119
- [Kuo et al. (2004)] Kuo, C. L., et al. 2004, *High-Resolution Observations of the Cosmic Microwave Background Power Spectrum with ACBAR*, *The Astrophysical Journal*, 600, 32
- [Lamarre et al. (2000)] Lamarre, J. M., et al. 2000, *The High Frequency Instrument of Planck: Design and Performances*, *Astrophysical Letters Communications*, 37, 161
- [Lamarre et al. (2003)] Lamarre, J. M., et al. 2003, *The Planck High Frequency Instrument, a third generation CMB experiment, and a full sky submillimeter survey*, *New Astronomy Review*, 47, 1017
- [Leitch et al. (1997)] Leitch, E. M., Readhead, A. C. S., Pearson, T. J., & Myers, S. T. 1997, *An Anomalous Component of Galactic Emission*, *Astrophysical Journal Letters*, 486, L23
- [Maino et al.(2002)] Maino, D., et al. 2002, *All-sky astrophysical component separation with Fast Independent Component Analysis (FASTICA)*, *MNRAS*, 334, 53
- [Maino et al.(2003)] Maino, D., Banday, A. J., Baccigalupi, C., Perrotta, F., & Górski, K. M. 2003, *Astrophysical component separation of COBE-DMR 4-yr data with FASTICA*, *MNRAS*, 344, 544
- [Maino et al.(2007)] Maino, D., Donzelli, S., Banday, A. J., Stivoli, F., & Baccigalupi, C. 2007, *Cosmic microwave background signal in Wilkinson Microwave Anisotropy Probe three-year data with FASTICA* *MNRAS*, 374, 1207

- [Mandolesi et al. (2000)] Mandolesi, N., Bersanelli, M., Burigana, C., & Villa, F. 2000, *The Planck Low Frequency Instrument*, Astrophysical Letters Communications, 37, 151
- [Maris et al. (2006)] Maris, M., Burigana, C., & Fogliani, S. 2006, *Astronomy and Astrophysics*, 452, 685
- [Martínez-González et al.(2003)] Martínez-González, E., Diego, J. M., Vielva, P., & Silk, J. 2003, *Cosmic microwave background power spectrum estimation and map reconstruction with the expectation-maximization algorithm*, MNRAS, 345, 1101
- [McCullough et al. (1999)] McCullough, P. R., & et al. 1999, *Implications of Halpha Observations for Studies of the CMB*, ASP Conf. Ser. 181: Microwave Foregrounds, 181, 253
- [Moudden et al. (2005)] Moudden, Y., Cardoso, J.-F., Starck, J.-L., Delabrouille, J. 2005, *Blind Component Separation in Wavelet Space: Application to CMB Analysis*, Eurasip Journal on Applied Signal Processing , 2005, 15 pp 2437-2454
- [Netterfield et al. (2002)] Netterfield, C. B., et al. 2002, *A Measurement by BOOMERANG of Multiple Peaks in the Angular Power Spectrum of the Cosmic Microwave Background*, The Astrophysical Journal, 571, 604
- [Patanchon et al. (2005)] Patanchon, G., Cardoso, J.-F., Delabrouille, J., & Vielva, P. 2005, *Cosmic microwave background and foregrounds in Wilkinson Microwave Anisotropy Probe first-year data*, MNRAS, 364, 1185
- [Ponthieu et al. (2005)] Ponthieu, N., et al. 2005, *Temperature and polarization angular power spectra of Galactic dust radiation at 353 GHz as measured by Archeops*, Astronomy and Astrophysics, 444, 327
- [Puget et al. (1996)] Puget, J.-L., Abergel, A., Bernard, J.-P., Boulanger, F., Burton, W. B., Desert, F.-X., & Hartmann, D. 1996, *Tentative detection of a cosmic far-infrared background with COBE*, Astronomy and Astrophysics, 308, L5
- [Readhead et al. (2004)] Readhead, A. C. S., et al. 2004, *Extended Mosaic Observations with the Cosmic Background Imager*, The Astrophysical Journal, 609, 498
- [Rephaeli (2002)] Rephaeli, Y. 2002, *The Sunyaev-Zeldovich effect: Recent progress and future prospects*, AIP Conf. Proc. 616: Experimental Cosmology at Millimetre Wavelengths, 616, 309
- [Sazonov & Sunyaev (1999)] Sazonov, S. Y., & Sunyaev, R. A. 1999, *Microwave polarization in the direction of galaxy clusters induced by the CMB quadrupole anisotropy* MNRAS, 310, 765
- [Shannon (1948)] Shannon, 1948, *A Mathematical Theory of Communication* Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July, October, 1948
- [Smoot (1998)] Smoot, G. F. 1998, *emphGalactic Free-Free and H-alpha emission*, ArXiv Astrophysics e-prints, arXiv:astro-ph/9801121
- [Stolyarov et al. (2002)] Stolyarov, V., Hobson, M. P., Ashdown, M. A. J., & Lasenby, A. N. 2002, *All-sky component separation for the Planck mission*, MNRAS, 336, 97
- [Sunyaev & Zeldovich (1972)] Sunyaev, R. A., & Zeldovich, Y. B. 1972, *The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies* Comments on Astrophysics and Space Physics, 4, 173
- [Sunyaev & Zeldovich (1980)] Sunyaev, R. A., & Zeldovich, I. B. 1980, *The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement* MNRAS, 190, 413

- [Tegmark & Efstathiou (1996)] Tegmark, M., & Efstathiou, G. 1996, *A method for subtracting foregrounds from multifrequency CMB sky maps*, MNRAS, 281, 1297
- [Tegmark et al.(2000)] Tegmark, M., Eisenstein, D. J., Hu, W., & de Oliveira-Costa, A. 2000, *Foregrounds and Forecasts for the Cosmic Microwave Background*, The Astrophysical Journal, 530, 133
- [Tegmark et al. (2003)] Tegmark, M., de Oliveira-Costa, A., & Hamilton, A. J. 2003, *High resolution foreground cleaned CMB map from WMAP* Phys. Rev. D., 68, 123523
- [Tristram et al. (2005)] Tristram, M., et al. 2005, *The CMB temperature power spectrum from an improved analysis of the Archeops data*, Astronomy and Astrophysics, 436, 785
- [Valls-Gabaud (1998)] Valls-Gabaud, D. 1998, *Cosmological applications of H-alpha surveys*, Publications of the Astronomical Society of Australia, 15, 111
- [White & Cohn (2002)] White, M., & Cohn, J. D. 2002, *The theory of anisotropies in the cosmic microwave background*, American Journal of Physics, 70, 106
- [Wiener (1949)] Wiener, N. 1949, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: Wiley. ISBN 0262730057
- [Yamada et al. (1999)] Yamada, M., Sugiyama, N., & Silk, J. 1999, *The Sunyaev-Zeldovich Effect by Cocoon of Radio Galaxies*, The Astrophysical Journal, 522, 66